Article:

Stochastic Galerkin finite volume shallow flow model: well-balanced treatment over uncertain topography

James Shaw, Ph.D[1] and Georges Kesserwani, Ph.D[2]

[1]Research Associate, Department of Civil and Structural Engineering, The University of Sheffield, Western Bank, Sheffield S10 2TN, U.K. Email: js102@zepler.net
[2]Research Fellow, Department of Civil and Structural Engineering, The University of Sheffield, Western Bank, Sheffield S10 2TN, U.K. Email: g.kesserwani@sheffield.ac.uk

ABSTRACT

Stochastic Galerkin methods can quantify uncertainty at a fraction of the computational expense of conventional Monte Carlo techniques, but such methods have rarely been studied for modelling shallow water flows. Existing stochastic shallow flow models are not well-balanced and their assessment has been limited to stochastic flows with smooth probability distributions. This paper addresses these limitations by formulating a one-dimensional stochastic Galerkin shallow flow model using a low-order Wiener-Hermite Polynomial Chaos expansion with a finite volume Godunov-type approach, incorporating the surface gradient method to guarantee well-balancing. Preservation of a lake-at-rest over uncertain topography is verified analytically and numerically. The model is also assessed using flows with discontinuous and highly non-Gaussian probability distributions. Prescribing constant inflow over uncertain topography, the model converges on a steady-state flow that is subcritical or transcritical depending on the topography elevation. Using only four Wiener-Hermite basis functions, the model produces probability distributions comparable to those from a Monte Carlo reference simulation with 2000 iterations, while executing about 100 times faster. Accompanying model software and simulation data is openly available online.

INTRODUCTION

Shallow water flows can be highly sensitive to uncertainties in input data: uncertainties in rating curve extrapolation, the choice of friction coefficient, digital elevation model errors and river survey measurement errors can substantially alter simulated distributions of flood extent, water depth and flow velocity (Bates et al., 2014; Kim and Sanders, 2016; Jung and Merwade, 2012; Montanari and Di Baldassarre, 2013). Conventional uncertainty quantification methods, including the Generalized Likelihood Uncertainty Estimation (GLUE) method (Beven and Binley, 1992), are based on Monte Carlo sampling with randomised inputs, making conventional methods computationally expensive and slow to converge on flow statistics. As a result, Monte Carlo simulations of flood events are severely constrained by available computing resources, and many sources of uncertainty must be neglected to make probabilistic simulations feasible (Neal et al., 2013). Stochastic Galerkin methods offer a computationally efficient alternative: repeated sampling is eliminated, and convergence is typically far more rapid than Monte Carlo methods (Xiu, 2009; Ge et al., 2008).

Polynomial Chaos methods work by introducing to the governing deterministic equations an additional, stochastic dimension, and approximating it with a so-called Polynomial Chaos expansion. Polynomial Chaos methods can be classified as non-intrusive or intrusive. Non-intrusive methods

Shaw, July 16, 2019

repeatedly sample a deterministic model with different input values, using the outputs to construct a stochastic solution. While non-intrusive polynomial chaos methods involve repeated sampling, they require far fewer samples than Monte Carlo methods (Ge et al., 2008; Ge and Cheung, 2011). Intrusive methods reformulate a deterministic model to produce a stochastic formulation that must be implemented as a new stochastic solver. The stochastic Galerkin method is an intrusive method which makes a Galerkin projection in stochastic space to produce a decoupled system of equations that are solved in a single model run to directly evolve the coefficients in the Polynomial Chaos expansion. Conventional stochastic Galerkin methods span the entire stochastic space using a single element with a basis chosen from the Askey scheme of orthogonal polynomials (Xiu and Karniadakis, 2002). When the stochastic flow is sufficiently smooth and the probability distributions are well-represented by the Polynomial Chaos basis, then Polynomial Chaos methods can achieve exponential convergence in the stochastic dimension (Xiu and Karniadakis, 2003). Sattar and El-Beltagy (2017) applied a conventional stochastic Galerkin method to the one-dimensional water hammer equations, and their model was able to converge on solutions with smooth, Gaussian probability distributions using a first-order Polynomial Chaos basis with just two basis functions.

Conventional stochastic Galerkin methods have two main shortcomings. First, strongly non-linear flows produce steep gradients or discontinuities in stochastic space that are often poorly represented by global Polynomial Chaos bases using a single element (Pettersson et al., 2014). Second, even when exponential convergence is achieved early in the simulation, solution accuracy can deteriorate severely over long simulations (Gerritsma et al., 2010). These shortcomings have motivated a wealth of research into better alternatives, including time-dependent basis functions (Gerritsma et al., 2010), multiresolution wavelet decompositions (Le Maître et al., 2004), multi-element discretisations and adaptive meshing of stochastic space (Wan and Karniadakis, 2006; Tryoen et al., 2010; Pettersson et al., 2014; Li and Stinis, 2015).

While stochastic Galerkin methods have received a great deal of attention, few have applied the method to the shallow water equations. Ge et al. (2008) made the first step in this direction, formulating a one-dimensional stochastic Galerkin shallow water model to study solitary wave propagation over uncertain topography. Their stochastic Galerkin model was 50 times more efficient than Monte Carlo with similar accuracy, demonstrating that stochastic Galerkin shallow water models offer a viable alternative to Monte Carlo methods. However, their numerical tests were limited to flows with smooth, unimodal probability distributions that are more easily represented by stochastic Galerkin methods. Furthermore, their stochastic Galerkin formulation used a centred difference approximation of the bed slope source term that is not well-balanced.

A well-balanced stochastic Galerkin model was formulated by Jin et al. (2016) and applied to scalar nonlinear equations including the inviscid Burgers' equation. The model inherited the well-balancing property from the underlying deterministic model that used the 'interface method' by Jin (2001). Jin et al. (2016) note that the intrusive stochastic Galerkin method has a particular advantage over non-intrusive stochastic collocation: a stochastic collocation method can only guarantee well-balancing at the collocation nodes where the deterministic well-balanced model is sampled. In contrast, the stochastic Galerkin model guarantees well-balancing across the entire stochastic space, irrespective of the order of Polynomial Chaos basis.

When applied to the shallow water equations, the well-balanced interface method is equivalent to the surface gradient method by Zhou et al. (2001). A stochastic reformulation of the surface gradient method is straightforward because it is a linear method involving only linear combinations of the discrete flow variables and topography. Popular alternatives to the surface gradient method include

     Shaw, July 16, 2019

well-balanced approaches that are also depth-positivity preserving (Audusse et al., 2004; Liang and Marche, 2009). Such approaches use nonlinear max($\cdot, \cdot$) operators which greatly complicate their stochastic Galerkin reformulation for realistic applications.

This paper presents a well-balanced stochastic Galerkin model of the one-dimensional shallow water equations for simulating probabilistic flows that account for measurement errors in flood plain topography or open channel bathymetry. This one-dimensional stochastic model represents the first step towards a two-dimensional probabilistic flood model, and it is also relevant to one-dimensional open channel flows. Being based on a Godunov-type shock-capturing method, the model is capable of simulating subcritical, supercritical and transcritical flows. The model uses a stochastic reformulation of the surface gradient method (Zhou et al., 2001), and the well-balancedness of this reformulation is studied theoretically and verified numerically for an idealised, motionless lake-at-rest over an uncertain and irregular bed. Next, a steady-state test is proposed that provokes a strongly nonlinear flow response over an uncertain bed, resulting in probability distributions that are highly non-Gaussian and discontinuous. Probability distributions produced by the stochastic Galerkin model are validated against a Monte Carlo reference simulation, and the relative computational performance is discussed. Finally, the stochastic Galerkin model is verified for a more realistic steady-state flow over a highly irregular bed with uncertainties that characterise measurement errors in bed elevation. Numerical simulation data (Shaw and Kesserwani, 2019a) and a Python 3 implementation of the stochastic Galerkin shallow flow model (Shaw and Kesserwani, 2019b) are available to download from Zenodo. Instructions for running the models and interpreting the data are provided in the appendix.

DETERMINISTIC AND STOCHASTIC SHALLOW FLOW MODELS

In this section, a certain deterministic numerical solver of the one dimensional (1D) shallow water equations is outlined in the framework of a finite volume Godunov-type method. The selected deterministic solver relies on the surface gradient method (Zhou et al., 2001) to both ensure a well-balanced topography integration and extendibility of the well-balanced property into the stochastic Galerkin case. Accordingly, a stochastic Galerkin reformulation is devised that is theoretically well-balanced with uncertain topography under a lake-at-rest hypothesis.

The mathematical model of the shallow water equations represent mass and momentum conservation principals, and is used in the following conservative form when solving it within a finite volume Godunov-type framework (Toro and Garcia-Navarro, 2007):

$$\frac{\partial \mathbf{U}(x,t)}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{U}(x,t))}{\partial x} = \mathbf{S}(\mathbf{U}(x,t), z(x)) \tag{1}$$

where $\mathbf{U} = [h, q]^\mathsf{T}$ is the flow vector including water depth h (L) and unit-width discharge $q = hv$ (L^2/T) in which v represents the depth-averaged velocity (L/T), $\mathbf{F} = \left[q, q^2/h + gh^2/2\right]^\mathsf{T}$ is the flux vector in which g represents the gravitational constant and $\mathbf{S} = [0, -gh \, dz/dx]^\mathsf{T}$ is the source term vector in which the gradient of the topography $z(x)$ is involved. Equation (1) represents hydraulic flow in an idealised, frictionless channel with a rectangular cross-section of unit width.

Deterministic model

On a uniform 1D mesh with M elements each of size Δx, the first-order finite volume method leads to the following discrete element-wise formulation of the shallow water equations given by

equation (1):

$$\mathbf{U}_i^{(n+1)} = \mathbf{U}_i^{(n)} - \Delta t \left(\frac{\tilde{\mathbf{F}}_{i+1/2}^{(n)} - \tilde{\mathbf{F}}_{i-1/2}^{(n)}}{\Delta x} - \mathbf{S}_i^{(n)} \right) \tag{2}$$

in which $\mathbf{U}_i^{(n)} = \left[h_i^{(n)}, q_i^{(n)} \right]^\top$ is a piecewise-constant discretisation of the flow vector at element i and time level (n), and $\tilde{\mathbf{F}}_{i+1/2}^{(n)}$ is a numerical flux function for linking nonlinear discontinuities associated with the flow vector data at interface $i+1/2$ located between element i and $i+1$. Namely, $\tilde{\mathbf{F}}_{i+1/2}^{(n)} = \tilde{\mathbf{F}}(\mathbf{U}_{i+1/2}^-, \mathbf{U}_{i+1/2}^+)$ where $\mathbf{U}_{i+1/2}^-$ is the limit of the solution from the side of element i, and $\mathbf{U}_{i+1/2}^+$ is the limit of the solution from the side of element $i+1$. Within the scope of this work involving a first-order accurate solver, these limits become $\mathbf{U}_{i+1/2}^- = \mathbf{U}_i$ and $\mathbf{U}_{i+1/2}^+ = \mathbf{U}_{i+1}$ that are used in a numerical flux function based on the Roe approximate Riemann solver (Roe and Pike, 1984). Consequently, the deterministic model is able to capture the occurrence of shocks, and simulate subcritical, supercritical and transcritical flows.

The surface gradient method essentially reconstructs an averaged topography at interface $i+1/2$ that is shared by both elements i and $i+1$, as $z_{i+1/2}^\star = (z_i + z_{i+1})/2$. From the reconstructed topography $z_{i+1/2}^\star$, consistent flow variable limits are accordingly reconstructed based on the actual free-surface elevation data, i.e. $\eta_{i+1/2}^- = h_i^{(n)} + z_i$ and $\eta_{i+1/2}^+ = h_{i+1}^{(n)} + z_{i+1}$, and velocity data, i.e. $v_{i+1/2}^- = q_i^{(n)}/h_i^{(n)}$ and $v_{i+1/2}^+ = q_{i+1}^{(n)}/h_{i+1}^{(n)}$, as: $h_{i+1/2}^{K,\star} = \eta_{i+1/2}^K - z_{i+1/2}^\star$ and $q_{i+1/2}^{K,\star} = h_{i+1/2}^{K,\star} v_{i+1/2}^K$ (where $K = +$ or $-$). For clarity of presentation, the time level denoted by superscript (n) is omitted from all reconstructed variables. These reconstructions form new Riemann states $\mathbf{U}_{i+1/2}^{K,\star} = \left[h_{i+1/2}^{K,\star}, q_{i+1/2}^{K,\star} \right]^\top$ for use to evaluate $\tilde{\mathbf{F}}_{i+1/2}^{(n)}$. By analogy, new Riemann limits $\mathbf{U}_{i-1/2}^{K,\star} = \left[h_{i-1/2}^{K,\star}, q_{i-1/2}^{K,\star} \right]^\top$ at $i-1/2$ are produced for use to evaluate $\tilde{\mathbf{F}}_{i-1/2}^{(n)}$. From the reconstructed limits, a well-balanced discretisation of the source term vector can be produced:

$$\mathbf{S}_i^{(n)} = \left[0, -g \left(\frac{h_{i-1/2}^{+,\star} + h_{i+1/2}^{-,\star}}{2} \right) \left(\frac{z_{i+1/2}^\star - z_{i-1/2}^\star}{\Delta x} \right) \right]^\top \tag{3}$$

The well-balanced deterministic model presented in equations (2) and (3) is used for Monte Carlo simulations, and the deterministic model is also the starting point for a stochastic Galerkin reformulation.

Fundamental properties of the Polynomial Chaos basis

Before presenting the stochastic Galerkin reformulation, it is necessary to consider a single random variable $A(\theta)$ that maps from the random event θ to an arbitrary probability distribution with finite variance. This random variable can be approximated by a Wiener-Hermite Polynomial Chaos expansion (Xiu and Karniadakis, 2002). The expansion is based on a standard Gaussian random variable $\xi(\theta) \in [-\infty, +\infty]$ having zero mean and unit variance. The random variable of interest, $A(\theta)$, is then approximated as

$$A(\theta) \approx \sum_{p=0}^{P} A_p \Phi_p(\xi(\theta)) \tag{4}$$

where $\mathbf{A} = [A_0, \ldots, A_P]^\mathsf{T}$ are the expansion coefficients and $\mathbf{\Phi} = [\Phi_0, \ldots, \Phi_P]^\mathsf{T}$ is the probabilists' Hermite polynomial basis having basis function Φ_p of degree p,

$$\Phi_p(\xi) = (-1)^p \exp\left(\frac{\xi^2}{2}\right) \frac{\mathrm{d}^p}{\mathrm{d}\xi^p} \exp\left(-\frac{\xi^2}{2}\right) \tag{5}$$

where $\Phi_0 = 1, \Phi_1 = \xi, \Phi_2 = \xi^2 - 1, \Phi_3 = \xi^3 - 3\xi$ and so on. As the basis order P is increased, the Wiener-Hermite Polynomial Chaos approximation converges on the true random variable $A(\theta)$ (Xiu and Karniadakis, 2002).

Basis orthogonality and commutativity

The Wiener-Hermite basis $\mathbf{\Phi}$ is orthogonal such that

$$\langle \Phi_p \Phi_s \rangle = \langle \Phi_p^2 \rangle \delta_{ps} \tag{6}$$

where $\langle \cdot \rangle$ is the ensemble average operator and δ_{ps} is the Kronecker delta that is equal to one when $p = s$ and zero otherwise. The ensemble average operator is defined as the weighted integral over the standard Gaussian random variable ξ:

$$\langle \alpha(\xi) \rangle = \int_{-\infty}^{\infty} \alpha(\xi) W(\xi) \, \mathrm{d}\xi \tag{7}$$

where $\alpha(\xi)$ is an expression involving any combination of random variables or basis functions, and the weighting function $W(\xi)$ is the standard Gaussian probability density function

$$W(\xi) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\xi^2}{2}\right) \tag{8}$$

This weighting function ensures that, when α is independent of ξ, the ensemble average $\langle \alpha \rangle = \alpha$. Finally, the ensemble average of a product of basis functions is commutative such that

$$\langle \Phi_p \Phi_s \rangle = \langle \Phi_s \Phi_p \rangle \tag{9}$$

The commutative property is needed later when verifying the well-balanced property with uncertain topography.

Stochastic Galerkin projection of a random variable

Given the random variable $A(\theta) = \sum_{p=0}^{P} A_p \Phi_p(\xi(\theta))$, its Galerkin projection onto a basis function Φ_l with $l = 0, \ldots, P$ is achieved using the ensemble average operator such that, due to orthogonality,

$$\langle A(\theta)\Phi_l \rangle = A_l \langle \Phi_l^2 \rangle \tag{10}$$

where A_l is the lth order expansion coefficient. Also note that the Galerkin projection of a basis function Φ_l and two random variables, $A(\theta)$ and $B(\theta)$, is distributive:

$$\langle (A(\theta) + B(\theta))\Phi_l \rangle = \langle A(\theta)\Phi_l \rangle + \langle B(\theta)\Phi_l \rangle \tag{11}$$

Mean, variance and high-order moments

The mean, variance and high-order moments can be calculated for $A(\theta)$. The mth moment $\mu_m[A]$ is defined as

$$\mu_m[A] = \int_{-\infty}^{\infty} (A - \beta)^m W(\xi) \, \mathrm{d}\xi = \langle (A - \beta)^m \rangle \tag{12}$$

where $\beta = 0$ when $m = 1$ and $\beta = \mu_1[A]$ for higher-order moments. Therefore, the mean $\mu_1[A] = \langle A \rangle = \sum_{p=0}^{P} A_p \langle \Phi_p \rangle$. Since $\langle \Phi_0 \rangle = 1$ and $\langle \Phi_p \rangle = 0$ for $p > 0$ then

$$\mu_1[A] = A_0 \tag{13}$$

The shorthand notation for the mean of A is $\overline{A}$, also known as the expected value, $\mathbb{E}[A]$.

The variance $\mu_2[X]$ can be derived using the fact that $\mathbb{E}\left[(A - \mathbb{E}[A])^2 \right] = \mathbb{E}[A^2] - \mathbb{E}^2[A]$, hence $\mu_2[A] = \left(\sum_{p=0}^{P} A_p^2 \langle \Phi_p^2 \rangle \right) - A_0^2 \langle \Phi_0 \rangle^2$. Since $\langle \Phi_0 \rangle^2 = \langle \Phi_0^2 \rangle$ then

$$\mu_2[A] = \sum_{p=1}^{P} A_p^2 \langle \Phi_p^2 \rangle \tag{14}$$

The shorthand notation for the variance of A is σ_A^2 and the standard deviation of A is σ_A.

Reconstructing the probability density function

The probability density function $f_A(a)$ of a random variable A is,

$$f_A(a) = \sum_{j=1}^{J} \left| \sum_{p=0}^{P} A_p \frac{\mathrm{d}\Phi_p}{\mathrm{d}\xi}(\hat{\xi}_j) \right|^{-1} W(\hat{\xi}_j) \tag{15a}$$

where $\hat{\xi}_j$, $j = 1, \ldots, J$ are the real roots of the polynomial

$$a - \sum_{p=0}^{P} A_p \Phi_p(\xi) = 0 \tag{15b}$$

which can be calculated numerically for a specific realisation a. Hence, the probability density function is computed by evaluating equation (15) for a range of outcomes.

Stochastic Galerkin reformulation of the deterministic model

The solution of the stochastic 1D shallow water equations is now random because it depends on uncertain initial conditions, uncertain boundary conditions and uncertain topography. Hence, the stochastic 1D shallow water equations depend not only upon space x and time t, but additionally upon the random event θ. The stochastic flow vector $\mathbf{U}(x, t, \theta)$ becomes a general stochastic process having arbitrary probability distributions that vary in space and time. Similarly, the stochastic topography $z(x, \theta)$ has arbitrary probability distributions that vary in space.

The stochastic Galerkin reformulation of the deterministic model involves three steps to (i) replace the deterministic variables, $\mathbf{U}_i^{(n)}$ and z_i, with random variables $\mathbf{U}_i^{(n)}(\theta)$ and $z_i(\theta)$, (ii) rewrite the deterministic formulation using these random variables, and (iii) make a stochastic Galerkin projection onto the Wiener-Hermite basis.

Shaw, July 16, 2019

Replacing deterministic variables with random variables

For all elements $i = 1, \ldots, M$ across all time levels, every deterministic flow variable $\mathbf{U}_i^{(n)} = \left[h_i^{(n)}, q_i^{(n)} \right]^\mathsf{T}$ and deterministic topography variable z_i becomes a random variable approximated by a Wiener-Hermite Polynomial Chaos expansion:

$$\mathbf{U}_i^{(n)}(\theta) \approx \sum_{p=0}^{P} \mathbf{U}_{i,p}^{(n)} \Phi_p(\xi(\theta)), \quad z_i(\theta) \approx \sum_{p=0}^{P} z_{i,p} \Phi_p(\xi(\theta)) \tag{16}$$

where $\mathbf{U}_{i,p}^{(n)} = \left[h_{i,p}^{(n)}, q_{i,p}^{(n)} \right]^\mathsf{T}$ and $z_{i,p}$ are the pth order expansion coefficients over element i at time level n.

The reconstructed topography and reconstructed limits become functions of random variables. The reconstructed topography at interface $i + 1/2$ becomes

$$\sum_{p=0}^{P} z_{i+1/2,p}^{\star} \Phi_p = \frac{1}{2} \left(\sum_{p=0}^{P} z_{i,p} \Phi_p + \sum_{p=0}^{P} z_{i+1,p} \Phi_p \right) \tag{17}$$

and so $z_{i+1/2,p}^{\star} = (z_{i,p} + z_{i+1,p})/2$ due to basis orthogonality. The reconstructed limits $\mathbf{U}_{i+1/2,p}^{K,\star} = \left[h_{i+1/2,p}^{K,\star}, q_{i+1/2,p}^{K,\star} \right]^\mathsf{T}$ (where $K = +$ or $-$) are calculated in a similar fashion.

Random variables that are functions of other random variables can be calculated in the same way. In particular, water depth can be expressed as a function of free-surface elevation and topography such that, due to basis orthogonality,

$$h_{i,p}^{(n)} = \eta_{i,p}^{(n)} - z_{i,p} \tag{18}$$

which is used later for specifying initial conditions.

Rewriting the deterministic formulation using random variables

The deterministic finite volume formulation given by equation (2) is rewritten in terms of the random variables in equation (16). As a result, the numerical fluxes $\tilde{\mathbf{F}}_{i+1/2}^{(n)}$ and $\tilde{\mathbf{F}}_{i-1/2}^{(n)}$ and source term vector $\mathbf{S}_i^{(n)}$ become functions of random variables. The numerical flux $\tilde{\mathbf{F}}_{i+1/2}^{(n)}$ becomes

$$\tilde{\mathbf{F}}_{i+1/2}^{(n)} = \tilde{\mathbf{F}} \left(\sum_{p=0}^{P} \mathbf{U}_{i+1/2,p}^{-,\star} \Phi_p, \sum_{p=0}^{P} \mathbf{U}_{i+1/2,p}^{+,\star} \Phi_p \right) \tag{19}$$

and similarly for $\tilde{\mathbf{F}}_{i-1/2}^{(n)}$. The source term vector $\mathbf{S}_i^{(n)}$ in equation (3) becomes

$$\mathbf{S}_i^{(n)} = \left[0, -\frac{g}{\Delta x} \left\{ \sum_{p=0}^{P} \left(\frac{h_{i-1/2,p}^{+,\star} + h_{i+1/2,p}^{-,\star}}{2} \right) \Phi_p \right\} \left\{ \sum_{s=0}^{P} \left(z_{i+1/2,s}^{\star} - z_{i-1/2,s}^{\star} \right) \Phi_s \right\} \right]^\mathsf{T} \tag{20}$$

Equation (20) involves the product of two expressions, each delimited by braces. Since both expressions include Wiener-Hermite expansions then different indices, p and s, are needed for the expansion coefficients in each expression.

Stochastic Galerkin projection

Due to the orthogonal property (equation 10) and distributive property (equation 11) of the Wiener-Hermite basis, a Galerkin projection of equation (2) onto the basis functions $\Phi_l, l = 0, \ldots, P$ produces $P + 1$ decoupled equations:

$$\mathbf{U}_{i,l}^{(n+1)} = \mathbf{U}_{i,l}^{(n)} - \frac{\Delta t}{\langle \Phi_l^2 \rangle} \left(\frac{\left\langle \tilde{\mathbf{F}}_{i+1/2}^{(n)} \Phi_l \right\rangle - \left\langle \tilde{\mathbf{F}}_{i-1/2}^{(n)} \Phi_l \right\rangle}{\Delta x} - \left\langle \mathbf{S}_i^{(n)} \Phi_l \right\rangle \right) \tag{21}$$

Equation (21) involves ensemble averages of numerical fluxes, $\left\langle \tilde{\mathbf{F}}_{i+1/2}^{(n)} \Phi_l \right\rangle$ and $\left\langle \tilde{\mathbf{F}}_{i-1/2}^{(n)} \Phi_l \right\rangle$, and an ensemble average of the source term vector, $\left\langle \mathbf{S}_i^{(n)} \Phi_l \right\rangle$. There is no straightforward method for calculating an ensemble average of the numerical flux because it is nonlinear. Instead, the integral in equation (7) is approximated by Gauss-Hermite quadrature,

$$\left\langle \tilde{\mathbf{F}}_{i+1/2}^{(n)} \Phi_l \right\rangle \approx \sum_{j=1}^{P+1} w_j \tilde{\mathbf{F}} \left(\sum_{p=0}^{P} \mathbf{U}_{i+1/2,p}^{-,\star} \Phi_p(\xi_j), \sum_{p=0}^{P} \mathbf{U}_{i+1/2,p}^{+,\star} \Phi_p(\xi_j) \right) \Phi_l(\xi_j) W(\xi_j) \tag{22}$$

where w_j are the quadrature weights and ξ_j are the quadrature points. The ensemble average $\left\langle \tilde{\mathbf{F}}_{i-1/2}^{(n)} \Phi_l \right\rangle$ is calculated in the same way.

Unlike the nonlinear numerical flux, the ensemble average of the source term vector $\left\langle \mathbf{S}_i^{(n)} \Phi_l \right\rangle$ is linear and can be derived directly from equation (20):

$$\left\langle \mathbf{S}_i^{(n)} \Phi_l \right\rangle = \left[0, -\frac{g}{\Delta x} \sum_{p=0}^{P} \sum_{s=0}^{P} \left(\frac{h_{i-1/2,p}^{+,\star} + h_{i+1/2,p}^{-,\star}}{2} \right) \left(z_{i+1/2,s}^{\star} - z_{i-1/2,s}^{\star} \right) \left\langle \Phi_p \Phi_s \Phi_l \right\rangle \right]^{\mathsf{T}} \tag{23}$$

The ensemble averages $\left\langle \Phi_p \Phi_s \Phi_l \right\rangle$ in equation (23) and $\left\langle \Phi_l^2 \right\rangle$ in equation (21) can be calculated analytically or exactly by Gauss-Hermite quadrature. Since these calculations do not depend on the solution then they can be precomputed once and stored.

Well-balanced Property with Uncertain Topography

In this section it is shown that the stochastic Galerkin model satisfies the well-balanced property for a lake-at-rest with uncertain topography. A lake at rest is an idealised scenario where the free-surface is entirely flat and motionless, much like still water in a lake. Assuming a lake-at-rest for the stochastic model, the mean free-surface elevation is constant and the mean discharge is zero, and there is no uncertainty in the free-surface elevation or discharge. The bed elevation can have any spatial profile and stochastic profile.

Since the free-surface elevation is constant and the averaged bed elevation is continuous at interfaces then $\eta_{i+1/2,p}^{-} - z_{i+1/2,p}^{\star} = \eta_{i+1/2,p}^{+} - z_{i+1/2,p}^{\star}$, hence $h_{i+1/2,p}^{-,\star} = h_{i+1/2,p}^{+,\star}$ for all $i = 0, \ldots, M$ and $p = 0, \ldots, P$. Shorthand notation is introduced such that η_p is the pth expansion coefficient of the spatially uniform free-surface elevation, and $h_{i+1/2,p}^{\star} = h_{i+1/2,p}^{-,\star} = h_{i+1/2,p}^{+,\star}$. Since $h_{i+1/2,p}^{\star}$ is continuous at interfaces and the discharge is zero then the numerical flux is equal to the physical flux.

Shaw, July 16, 2019

In order to preserve a lake-at-rest solution in element i then the ensemble average of the flux gradient must balance the ensemble average of the source term vector in equation (21). The balance of mass and momentum components can be considered separately. The mass continuity equation balances because the discharge is zero. For the momentum equation, the following equality must hold for all $l = 0, \ldots, P$:

$$\frac{\left\langle F^{(n)}_{i+1/2}\Phi_l \right\rangle - \left\langle F^{(n)}_{i-1/2}\Phi_l \right\rangle}{\Delta x} - \left\langle S^{(n)}_i \Phi_l \right\rangle = 0 \tag{24}$$

where the ensemble average of the bed slope source term $\left\langle S^{(n)}_i \Phi_l \right\rangle$ is given by equation (23) and the momentum flux $F^{(n)}_{i+1/2}$ is

$$F^{(n)}_{i+1/2} = \frac{g}{2}\left\{ \sum_{p=0}^{P} h^{\star}_{i+1/2,p}\Phi_p \right\}\left\{ \sum_{s=0}^{P} h^{\star}_{i+1/2,s}\Phi_s \right\} \tag{25}$$

and similarly for $F^{(n)}_{i-1/2}$. Due to the distributive property of the basis (equation 11), equation (24) can be rewritten as

$$\left\langle \frac{F^{(n)}_{i+1/2} - F^{(n)}_{i-1/2}}{\Delta x}\Phi_l \right\rangle = \left\langle S^{(n)}_i \Phi_l \right\rangle \tag{26}$$

The ensemble average of the momentum flux gradient is

$$\left\langle \frac{F^{(n)}_{i+1/2} - F^{(n)}_{i-1/2}}{\Delta x}\Phi_l \right\rangle = \frac{g}{2\Delta x}\sum_{p=0}^{P}\sum_{s=0}^{P}\left(h^{\star}_{i+1/2,p}\,h^{\star}_{i+1/2,s} - h^{\star}_{i-1/2,p}\,h^{\star}_{i-1/2,s} \right)\left\langle \Phi_p\Phi_s\Phi_l \right\rangle \tag{27}$$

To be able to factorise equation (27), the first step is to add then subtract the term $h^{\star}_{i-1/2,p}h^{\star}_{i+1/2,s}$ to yield

$$\left\langle \frac{F^{(n)}_{i+1/2} - F^{(n)}_{i-1/2}}{\Delta x}\Phi_l \right\rangle = \frac{g}{2\Delta x}\sum_{p=0}^{P}\sum_{s=0}^{P}\left(h^{\star}_{i+1/2,p}\,h^{\star}_{i+1/2,s} - h^{\star}_{i-1/2,p}\,h^{\star}_{i-1/2,s}+ \right.$$
$$\left. h^{\star}_{i-1/2,p}\,h^{\star}_{i+1/2,s} - h^{\star}_{i-1/2,p}\,h^{\star}_{i+1/2,s} \right)\left\langle \Phi_p\Phi_s\Phi_l \right\rangle \tag{28}$$

Using the distributive property of multiplication, and associativity of summation, equation (28) can be rewritten as

$$\left\langle \frac{F^{(n)}_{i+1/2} - F^{(n)}_{i-1/2}}{\Delta x}\Phi_l \right\rangle = \frac{g}{2\Delta x}\left[\sum_{p=0}^{P}\sum_{s=0}^{P}\left(h^{\star}_{i+1/2,p}\,h^{\star}_{i+1/2,s} - h^{\star}_{i-1/2,p}\,h^{\star}_{i-1/2,s} + h^{\star}_{i-1/2,p}\,h^{\star}_{i+1/2,s} \right)\left\langle \Phi_p\Phi_s\Phi_l \right\rangle \right.$$
$$\left. - \sum_{p=0}^{P}\sum_{s=0}^{P} h^{\star}_{i-1/2,p}\,h^{\star}_{i+1/2,s}\left\langle \Phi_p\Phi_s\Phi_l \right\rangle \right] \tag{29}$$

The final term in equation (29) can be rewritten as $\sum_{p=0}^{P}\sum_{s=0}^{P} h_{i-1/2,s}^{\star} h_{i+1/2,p}^{\star} \langle \Phi_p \Phi_s \Phi_l \rangle$ due to commutativity of summation and product operators as well as the basis function (equation 9). Then, by associativity of summation:

$$\left\langle \frac{F_{i+1/2}^{(n)} - F_{i-1/2}^{(n)}}{\Delta x} \Phi_l \right\rangle = \frac{g}{2\Delta x} \sum_{p=0}^{P} \sum_{s=0}^{P} \left(h_{i+1/2,p}^{\star} h_{i+1/2,s}^{\star} - h_{i-1/2,p}^{\star} h_{i-1/2,s}^{\star} + \right.$$
$$\left. h_{i-1/2,p}^{\star} h_{i+1/2,s}^{\star} - h_{i-1/2,s}^{\star} h_{i+1/2,p}^{\star} \right) \langle \Phi_p \Phi_s \Phi_l \rangle \quad (30)$$

Finally, equation (30) can be factorised as

$$\left\langle \frac{F_{i+1/2}^{(n)} - F_{i-1/2}^{(n)}}{\Delta x} \Phi_l \right\rangle = \frac{g}{2\Delta x} \sum_{p=0}^{P} \sum_{s=0}^{P} \left(h_{i-1/2,p}^{\star} + h_{i+1/2,p}^{\star} \right) \left(h_{i+1/2,s}^{\star} - h_{i-1/2,s}^{\star} \right) \langle \Phi_p \Phi_s \Phi_l \rangle \quad (31)$$

and, given that $h_{i+1/2,s}^{\star} = \eta_s - z_{i+1/2,s}^{\star}$ and $h_{i-1/2,s}^{\star} = \eta_s - z_{i-1/2,s}^{\star}$, then

$$\left\langle \frac{F_{i+1/2}^{(n)} - F_{i-1/2}^{(n)}}{\Delta x} \Phi_l \right\rangle = -\frac{g}{\Delta x} \sum_{p=0}^{P} \sum_{s=0}^{P} \left(\frac{h_{i-1/2,p}^{+,\star} + h_{i+1/2,p}^{-,\star}}{2} \right) \left(z_{i+1/2,s}^{\star} - z_{i-1/2,s}^{\star} \right) \langle \Phi_p \Phi_s \Phi_l \rangle \quad (32)$$

Equation (32) is equal to the ensemble average of the bed slope source term $\left\langle S_i^{(n)} \Phi_l \right\rangle$ given in equation (23), hence discrete balance is preserved.

IDEALISED NUMERICAL TESTS

Three numerical tests are performed to validate the well-balanced stochastic Galerkin model. The first test simulates a lake-at-rest over idealised, uncertain topography to verify that the stochastic Galerkin model preserves well-balancing numerically. The second test simulates flow over a hump with an uncertain elevation. This test is designed to challenge the stochastic Galerkin model at representing discontinuous, non-Gaussian probability distributions by generating a steady-state solution that may be subcritical or transcritical depending on the hump elevation. The third test verifies the robustness of the stochastic Galerkin model for a steady flow over a highly irregular and uncertain bed that is more representative of real-world river hydraulics.

Specification of Tests 1 and 2

For the first two tests, the 1D domain is $[-50\,\mathrm{m}, 50\,\mathrm{m}]$, tessellated by $M = 100$ elements with no overlaps or gaps such that the mesh spacing is $\Delta x = 1\,\mathrm{m}$. The timestep is $\Delta t = 0.15\,\mathrm{s}$ resulting in a maximum Courant number of about 0.8. By choosing a fixed timestep, simulations of a given test complete in the same number of timesteps irrespective of the model configuration, and error accumulation due to timestepping errors will be the same across all models.

Both tests include a topographic hump centred at $x = 0\,\mathrm{m}$ with a region of Gaussian uncertainty. Following a similar approach to Ge et al. (2008), there are two representations for the same uncertain topography. The first representation enables smooth topography profiles to be randomly generated in Monte Carlo iterations. The topography z is defined as

$$z(x, r) = r \operatorname{sech}^2 \left(\frac{\pi x}{\lambda} \right) \quad (33)$$

where the hump amplitude r is a random variable with mean $\bar{r} = 0.6\,\text{m}$ and standard deviation $\sigma_r = 0.3\,\text{m}$, and the half-width is $\lambda = 10\,\text{m}$. This topography profile is seen in figure 3. The second representation is used by the stochastic Galerkin model, with topography represented by expansion coefficients $z_{i,0}, \ldots, z_{i,P}$. To be able to calculate the topography expansion coefficients, equation (33) must be expressed in terms of mean topography $\bar{z}(x)$ and topographic variance $\sigma_z^2(x)$ without involving the random variable r. The mean topography $\bar{z}(x)$ is simply

$$\bar{z}(x) = z(x, \bar{r}) \tag{34}$$

The topographic variance is

$$\sigma_z^2(x) = \mathbb{E}[z^2(x, r)] - \mathbb{E}^2[z(x, r)] \tag{35}$$

Equation (35) can be rewritten using Taylor series expansions of the two terms $\mathbb{E}\left[z^2(x, r)\right]$ and $\mathbb{E}^2\left[z(x, r)\right]$. To illustrate the approach, a Taylor series expansion of $\mathbb{E}\left[z(x, r)\right]$ about $\bar{r}$ is

$$\mathbb{E}\left[z(x, r)\right] = \mathbb{E}\left[z(x, \bar{r} + (r - \bar{r}))\right]$$
$$= \mathbb{E}\left[z + \frac{\partial z}{\partial r}(r - \bar{r}) + \frac{1}{2}\frac{\partial^2 z}{\partial r^2}(r - \bar{r})^2 + O(r^3)\right] \tag{36}$$

where z is shorthand for $z(x, \bar{r})$ and $O(r^3)$ is the error term involving high-order derivatives $\partial^m z / \partial r^m$ with $m \geq 3$. Since $\mathbb{E}\left[r - \bar{r}\right] = 0$ and $\mathbb{E}\left[(r - \bar{r})^2\right] = \sigma_r^2$ then equation (36) simplifies to

$$\mathbb{E}\left[z(x, r)\right] = z + \frac{1}{2}\frac{\partial^2 z}{\partial r^2}\sigma_r^2 + O(r^3) \tag{37}$$

Applying this approach to equation (35) gives:

$$\sigma_z^2(x) = z^2 + \left[\left(\frac{\partial z}{\partial r}\right)^2 + z\frac{\partial^2 z}{\partial r^2}\right]\sigma_r^2 - \left[z + \frac{1}{2}\frac{\partial^2 z}{\partial r^2}\sigma_r^2\right]^2 + O(r^3) \tag{38}$$

For the topographic profile given by equation (33), it holds that $\partial^m z / \partial r^m = 0$ where $m \geq 2$, so the Taylor series approximation introduces no spurious oscillations in stochastic space. The topographic variance in equation (38) then simplifies to

$$\sigma_z^2(x) = \left(\frac{\partial z(x, \bar{r})}{\partial r}\sigma_r\right)^2 \tag{39}$$

Equipped with analytic expressions for the mean topography $\bar{z}(x)$ (equation 34) and topographic variance $\sigma_z^2(x)$ (equation 39), now the topography expansion coefficients $z_{i,0}, \ldots, z_{i,P}$ can be calculated. Since the topographic bump has a Gaussian probability distribution with $\mu_1[z(x)] = \bar{z}(x)$, $\mu_2[z(x)] = \sigma_z^2(x)$ and high-order moments $\mu_m[z(x)] = 0$ for $m \geq 3$ then, using equations (12), (13) and (14), the topography expansion coefficients are

$$z_{i,p} = \begin{cases} \bar{z}(x_i) & \text{if } p = 0 \\ \sigma_z(x_i) & \text{if } p = 1 \\ 0 & \text{otherwise} \end{cases} \tag{40}$$

where values are calculated at the centre point x_i for all elements $i = 1, \ldots, M$.

The initial water depth expansion coefficients $h_{i,0}^{(0)}, \ldots, h_{i,P}^{(0)}$ can be calculated in terms of free-surface elevation and topography using equation (18). For both tests, the initial, spatially-uniform mean free-surface elevation is 1.5 m with no initial uncertainty such that $\eta_{i,0}^{(0)} = 1.5$ m and $\eta_{i,p}^{(0)} = 0$ with $p = 1, \ldots, P$ and $i = 1, \ldots, M$.

Test 1: Lake-at-Rest Over an Uncertain Bed

Numerical methods that are not well-balanced produce spurious waves in the vicinity of sloping topography, and these spurious waves are particularly evident for slow-moving flows with weak momentum fluxes. In the limit, the momentum flux is zero and the water is motionless, like the water surface on a lake at rest. Hence, the lake-at-rest test is ideally suited to verify the well-balanced property, since the analytic solution preserves the resting state forever, and any waves generated by a numerical model are entirely spurious (Bermúdez and Vázquez, 1994).

To present a challenging test, a rectangular obstacle is introduced to the right of the uncertain hump given by equation (33), so the bed elevation z becomes

$$z(x, r) = r \operatorname{sech}^2 \left(\frac{\pi x}{\lambda} \right) + \check{z}(x), \tag{41a}$$

where $\check{z}$ is the rectangular obstacle:

$$\check{z}(x) = \begin{cases} \bar{r} & \text{if } 30 < x \le 40, \\ 0 & \text{otherwise} \end{cases} \tag{41b}$$

Results of the well-balanced stochastic model are compared with those of a stochastic model having a centred difference approximation of the source term vector that does not exactly balance the numerical flux gradient. The centred difference model is the same as the well-balanced stochastic model except for two changes. First, numerical fluxes are calculated using the original, unmodified flow variables:

$$\tilde{\mathbf{F}}_{i+1/2}^{(n)} = \tilde{\mathbf{F}} \left(\sum_{p=0}^{P} \mathbf{U}_{i+1/2,p}^{-} \Phi_p, \sum_{p=0}^{P} \mathbf{U}_{i+1/2,p}^{+} \Phi_p \right) \tag{42}$$

Second, the ensemble average of the source term vector uses a centred difference approximation:

$$\langle \mathbf{S}_i \Phi_l \rangle = \left[0, -g \sum_{p=0}^{P} \sum_{s=0}^{P} h_{i,p} \frac{z_{i+1,s} - z_{i-1,s}}{2\Delta x} \langle \Phi_p \Phi_s \Phi_l \rangle \right]^{\mathsf{T}} \tag{43}$$

The centred difference model and the well-balanced model are both configured with a Wiener-Hermite basis order $P = 3$.

The simulated time is 100 s corresponding to about 670 timesteps, and the solutions for the centred difference and well-balanced stochastic Galerkin models are shown in figure 1. The lack of well-balancing is apparent using the centred difference model: grid-scale standing waves develop at the discontinuities either side of the rectangular hump (figure 1a, 1c), and a smooth standing wave also develops over the uncertain hump. These errors persist throughout the simulation. In contrast, the well-balanced stochastic Galerkin model preserves the initial resting state with
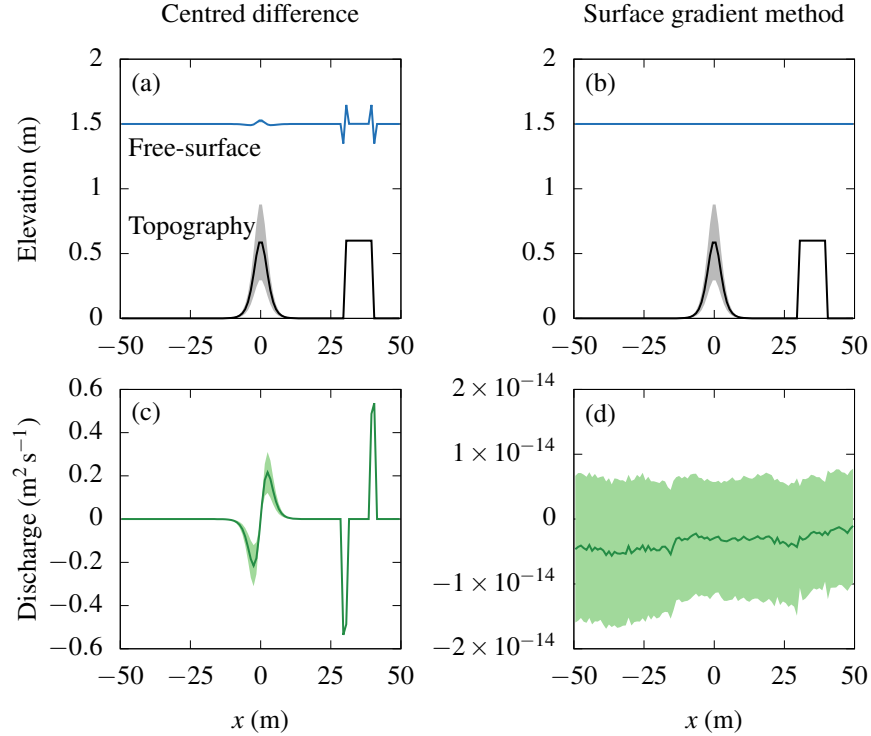
12

Fig. 1. Stochastic lake-at-rest solutions at $t = 100$ s. Mean values are marked by solid lines and shaded regions represent one standard deviation from the mean.

discharges accurate to machine precision (figure 1d). This numerical result confirms that the stochastic Galerkin model is well-balanced in theory and in practice.

The choice of the Wiener-Hermite basis introduces a particular limitation that imposes an upper bound on the basis order P, and constrains the minimum water depth that the stochastic Galerkin model can represent. This limitation arises because the hump amplitude has a Gaussian probability distribution so the tails of the distribution extend to $\pm\infty$, meaning that there is a non-zero probability that the water depth is negative. The stochastic Galerkin formulation presented here does not accommodate wetting-and-drying processes, and any negative water depth will crash the model. If the basis order P is increased then the Gauss-Hermite quadrature points in equation (22) extend further into the tails of the probability distributions, leading to negative water depths being provided as input to the Riemann solver. Similarly, raising the topography, decreasing the initial water depth, or increasing the topographic uncertainty can all produce negative water depths in the stochastic Galerkin model. This behaviour has been verified experimentally by varying the model basis order, initial conditions and topography profile.

Test 2: Steady-State Critical Flow Over an Uncertain Bed

This test is designed to challenge the stochastic Galerkin model at representing highly non-Gaussian and discontinuous distributions of stochastic flow. The uncertain topography profile and inflow boundary condition are chosen in order to produce a nonlinear response such that the steady-state solution may be subcritical or transcritical depending on the bed elevation. Results of the stochastic Galerkin model are validated against a Monte Carlo simulation that serves as a
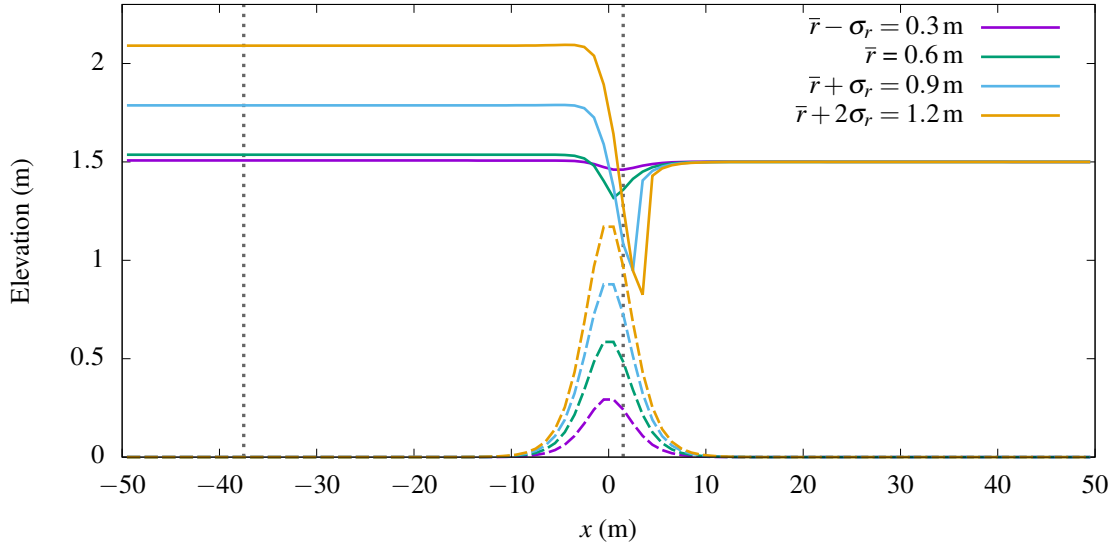
Shaw, July 16, 2019

Fig. 2. Well-balanced deterministic solutions of steady-state flow at $t = 500\,$s using four hump amplitudes, $\bar{r} - \sigma_r = 0.3\,$m, $\bar{r} = 0.6\,$m, $\bar{r} + \sigma_a = 0.9\,$m and $\bar{r} + 2\sigma_r = 1.2\,$m. The free-surface elevation is shown with solid lines and the topography profile is shown with dashed lines. Vertical dotted lines at $x = -37.5\,$m and $x = 1.5\,$m mark the positions of the probability densities shown in figure 4.

reference solution.

Setup to Produce a Nonlinear Flow Response

The uncertain topography is the smooth hump given by equation (40). Subcritical boundary conditions are imposed such that the mean upstream discharge per unit-width is $1.65\,\mathrm{m^2\,s^{-1}}$ and the mean downstream water depth is $1.5\,$m, with no uncertainty for the upstream discharge and downstream water depth. Transmissive boundary conditions are used for the upstream water depth and downstream discharge. These boundary conditions are chosen so that the flow is exactly critical for the mean hump amplitude $\bar{r} = 0.6\,$m at $x = 0\,$m. Since the hump amplitude is uncertain then the flow regime is also uncertain: if the hump amplitude is less than $\bar{r}$ then the flow remains subcritical; if the hump amplitude is greater than $\bar{r}$ then the flow regime becomes transcritical. In the transcritical regime, the flow upstream of the hump is subcritical, transitioning to supercritical flow over the hump. A hydraulic jump occurs on the downstream side of the hump, where the flow becomes subcritical once more.

To illustrate this change in flow regime, figure 2 shows four deterministic solutions using four different hump amplitudes. Solutions from the well-balanced deterministic model are obtained at $t = 500\,$s when the water has converged on a steady state. Convergence is measured by calculating the L^2 difference in mean water depth between the current and previous timesteps:

$$L^2 \text{ difference in mean water depth} = \sqrt{\sum_{i=1}^{M} \left(h_{i,0}^{(n)} - h_{i,0}^{(n-1)} \right)^2} \tag{44}$$

By $t = 500$ s all four deterministic solutions have converged down to a convergence error of 10^{-4} m. For a small hump with amplitude $\bar{r} - \sigma_r = 0.3$ m, the flow remains subcritical. A linear increase in hump amplitude produces a strongly nonlinear response in the steady-state water profile, as seen in figure 2. Two nonlinear responses are evident in particular: first, the upstream boundary condition allows the upstream water depth to increase nonlinearly; second, a transcritical shock develops over the hump that increases in amplitude and moves further downstream with larger hump amplitudes. Downstream of the hump, the water depth is 1.5 m irrespective of the hump amplitude, with this profile having propagated upstream from the imposed downstream boundary.

Configuration of the Monte Carlo Reference Simulation

The Monte Carlo reference simulation is performed using iterations of the well-balanced deterministic model. It is necessary to perform a sufficient number of Monte Carlo iterations to ensure that the flow statistics are accurate, and more iterations are needed for more complex probability distributions. After each Monte Carlo iteration, the mean and standard deviation of water depth at $x = 1.5$ m are measured, where the probability distribution is most complex. The statistics are compared with those of the previous iteration, and the Monte Carlo simulation terminates once the statistics change negligibly between iterations, when statistical convergence is achieved. Guided by these measurements, two thousand Monte Carlo iterations are needed to achieve statistical convergence for this test.

For each Monte Carlo iteration, the topography is randomly generated using a hump amplitude drawn from the Gaussian distribution given by $(\bar{r}, \sigma_r)$ and so the topography will always be smooth. If instead the topography was randomly generated using $(\bar{z}(x), \sigma_z(x))$ then randomisation would be different in every element and the topography would not be smooth, so many more iterations would be needed to sample the stochastic solution space. For the Monte Carlo iterations, the hump amplitude r is constrained such that 0 m $\leq r \leq 1.4$ m to avoid negative water depths.

Spatial Profiles of the Uncertain Free-Surface Elevation

In figure 3, spatial profiles of the uncertain free-surface elevation are obtained at $t = 500$ s when the water depth profiles from the Monte Carlo and stochastic Galerkin simulations have converged down to 10^{-4} m as defined by equation (44). Using a Wiener-Hermite basis of order $P = 3$, the stochastic Galerkin model accurately represents the Monte Carlo reference profiles of the mean and standard deviation of free-surface elevation. Upstream of the hump, the stochastic Galerkin model predicts a standard deviation that is slightly too small compared to the Monte Carlo reference solution. Small errors in the stochastic Galerkin free-surface elevation are also visible above the hump where the flow is most complex. Similar results are obtained using basis order $P = 1$ or $P = 2$, with stochastic Galerkin errors increasing slightly as the basis order is decreased (not shown).

Monte Carlo Histograms of Uncertain Free-Surface Elevation

The mean and standard deviation statistics are useful for summarising the spatial profile of uncertainty, but they are less meaningful for non-Gaussian probability distributions. In such cases, it is more meaningful to study the complete probability distributions, which is also particularly important for flood risk assessments that are concerned with extreme events that occur in the tails of the distributions (Ge and Cheung, 2011).

Probability distributions of the free-surface elevation are sampled at two points at $t = 500$ s: the first at $x = -37.5$ m and the second at $x = 1.5$ m, with these positions marked by dotted lines in
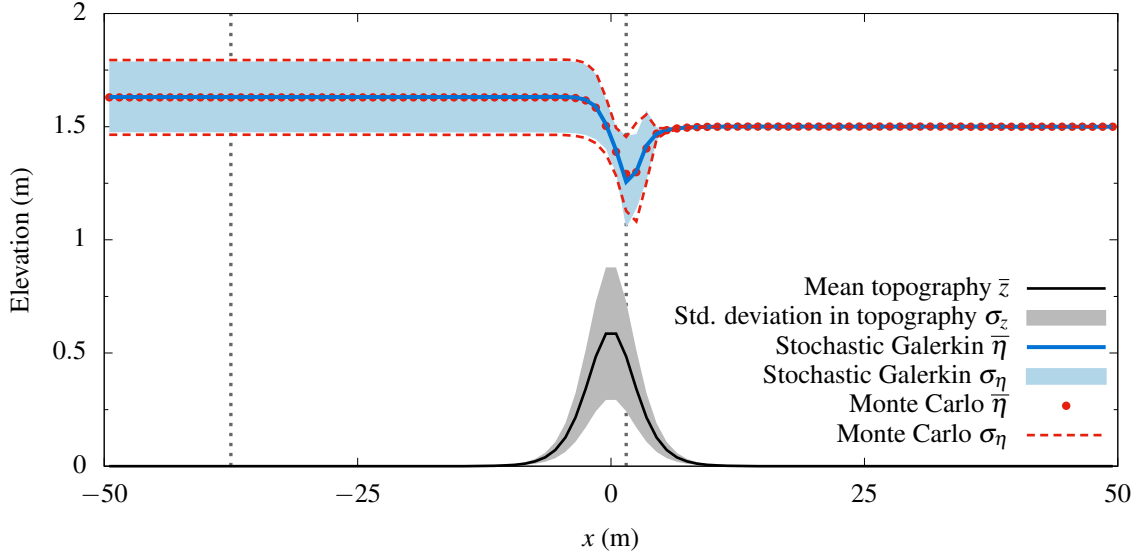
Fig. 3. Solutions of steady state critical flow over an uncertain hump at $t = 500$ s, comparing stochastic Galerkin and Monte Carlo profiles of mean free-surface elevation $\overline{\eta}$ and standard deviation of free-surface elevation σ_η. The stochastic Galerkin result is obtained using basis order $P = 3$. Vertical dotted lines at $x = -37.5$ m and $x = 1.5$ m mark the positions of the probability densities shown in figure 4.
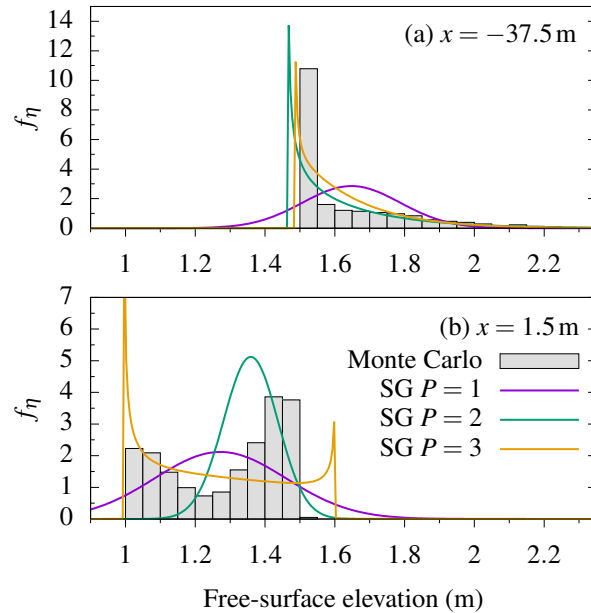

Fig. 4. Probability distributions of free-surface elevation, f_η, at (a) $x = -37.5$ m and (b) $x = 1.5$ m for steady state critical flow over an uncertain hump at $t = 500$ s. For the Monte Carlo reference simulation, probability distributions are estimated by histograms. Continuous probability density functions are reconstructed from stochastic Galerkin (SG) results using basis orders $P = 1$, $P = 2$ and $P = 3$.

Shaw, July 16, 2019

figure 2 and figure 3. The first point is far upstream of the hump where the free-surface elevation is uncertain and spatially uniform. The second point is immediately downstream of the hump in the region where transcritical shocks can develop.

Figure 4 shows histograms from the Monte Carlo reference simulation that estimate the true probability densities at the two points. Stochastic Galerkin results which appear on the same figure are discussed later. For each of the two points, water depths from the 2000 Monte Carlo iterations are binned into intervals of 0.05 m, and the magnitude of each bin represents the probability that the water depth is within the given interval. Since the histogram estimates the probability density then the total shaded area over all bins is equal to one.

The Monte Carlo histogram at $x = -37.5$ m is shown in figure 4a and is discussed first. For subcritical flows over small humps, the upstream water level remains at its initial height of 1.5 m. For transcritical flows over larger humps, the upstream water level increases nonlinearly. Since the initial conditions and boundary conditions are chosen so that the mean flow is critical, then about 50% of the flows are subcritical, resulting in a large peak in the probability distribution at $\eta = 1.5$ m. The other 50% of the flows are transcritical with elevated upstream water levels, resulting in a long tail in the distribution.

The stochastic flow response immediately downstream of the hump at $x = 1.5$ m is more complex. For subcritical flows, the steady-state free-surface elevation at this point will be only slightly lower than the initial free-surface elevation of 1.5 m. For transcritical flows, the steady-state free-surface elevation may be much lower since, at $x = 1.5$ m, the transcritical shock is close to its minimum depth. The subcritical and transcritical regimes appear as a bimodal distribution in the histogram in figure 4b with one peak around $\eta = 1.0$ m associated with transcritical flow and a second peak around $\eta = 1.5$ m associated with subcritical flow.

Stochastic Galerkin Free-Surface Elevation Probability Densities

Overlayed on the Monte Carlo histograms, figure 4 also shows probability density functions obtained from stochastic Galerkin simulations. Three stochastic Galerkin simulations are performed using Wiener-Hermite bases of order $P = 1$, $P = 2$ and $P = 3$. Free-surface elevation expansion coefficients are calculated by rearranging equation (18), from which probability density functions are reconstructed using equation (15). If a probability density function was in exact agreement with a Monte Carlo histogram then the line would pass through the top of every histogram bin, and so any deviation represents a numerical error associated with the stochastic Galerkin model.

Using basis order $P = 1$, the stochastic Galerkin model can only represent Gaussian distributions, with its two expansion coefficients representing the mean and standard deviation. While figure 3 confirms that the mean and standard deviation statistics from the stochastic Galerkin model are accurate, the Monte Carlo histograms cannot be well-represented by Gaussian distributions.

Increasing the basis order to $P = 2$, the probability distribution of upstream water levels is in good agreement with the Monte Carlo histogram (figure 4a), though an error is noticeable at the discontinuity: the Monte Carlo histogram has a discontinuity at $\eta = 1.5$ m because the upstream water level can only rise above its initial height, so the probability that $\eta < 1.5$ m is zero. The stochastic Galerkin model with basis order $P = 2$ slightly underestimates this discontinuity, placing it at about $\eta = 1.45$ m. The distribution in free-surface elevation at $x = 1.5$ m (figure 4b) is not improved using basis order $P = 2$, with the distribution remaining close to Gaussian.

Finally, the basis order is increased to $P = 3$. The distribution of upstream water levels shifts slightly to the right in figure 4a, corresponding to slightly higher water levels. As a result, the

Shaw, July 16, 2019

discontinuity is closer to the true value of $\eta = 1.5$ m, but this shift also produces slightly larger errors in the tail of the distribution. The most notable improvement is seen in the distribution at $x = 1.5$ m (figure 4b): the distribution from the stochastic Galerkin model now has a similar shape to the Monte Carlo histogram. The lower bound at $\eta = 1.0$ m is accurately represented, but the true upper bound around $\eta = 1.5$ m is overestimated by about 0.1 m. The probability density function has singularities at the lower and upper bounds that appear as spikes in the plot. These two singularities occur because the probability density function is the derivative of the cumulative density function, which is discontinuous and non-differentiable at these points.

While the stochastic Galerkin model with basis order $P = 3$ adequately represented the true distribution bounds and distribution shapes, the probability density functions were not entirely accurate. Such inaccuracies are to be expected because low-order Wiener-Hermite bases cannot represent complex distributions. The basis order cannot be increased beyond $P = 3$ because water depths are small near transcritical shocks, and the stochastic Galerkin model crashes for the same reason as discussed in the lake-at-rest test. Instead, results might be improved by choosing a more sophisticated method to discretise stochastic space. One candidate is the stochastic Galerkin multiwavelet approach by Pettersson et al. (2014), which is able to simulate stochastic gas dynamics with densities close to zero, analogous to very small water depths in shallow water flows.

Storage Requirements and Computation Time

Since the stochastic Galerkin model is necessarily more complex than the deterministic model, associated increases in storage requirements and computation time are expected. Stochastic Galerkin storage requirements scale linearly with the chosen basis order P because the model stores $P + 1$ expansion coefficients per variable per element. The ensemble averages $\langle \Phi_p \Phi_s \Phi_l \rangle$ in equation (23) and $\langle \Phi_l^2 \rangle$ in equation (21) are constant values that can be precomputed once and stored. Due to the commutative property of the ensemble average of basis function products (equation 9), the associated storage requirements are small.

The expected increase in computation time can be estimated by examining the model formulation and assuming that calculations are performed without parallelisation. Choosing basis order P, the entire evolution equation (equation 21) must be evaluated $P + 1$ times for the basis functions $0, \ldots, P$. For each evaluation, the ensemble average of numerical fluxes is calculated by sampling the Riemann solver $P + 1$ times using Gauss-Hermite quadrature (equation 22). The computation time for the ensemble average of the source term vector (equation 23) can be neglected because its calculation is trivially fast compared to the Riemann solver. Hence, by considering the total samples of the Riemann solver, it is expected that the stochastic Galerkin model will be at least $(P + 1)^2$ times slower than the deterministic solver.

Measuring the elapsed CPU time to simulate the steady-state test confirms this expectation: the stochastic Galerkin model with basis order $P = 3$ is in fact about 20 times slower than the deterministic model, and the Riemann solver accounts for about 90% of the stochastic Galerkin total computation time. Compared to the Monte Carlo simulation which uses 2000 deterministic iterations, the stochastic Galerkin model is about 100 times faster. Note that the deterministic model, stochastic Galerkin model and Monte Carlo simulation were implemented without parallelisation.

The speed-up observed here compares favourably with numerical tests performed by Ge et al. (2008). Their stochastic Galerkin model with basis order $P = 5$ was about 200 times slower than their deterministic model. Their Monte Carlo simulation used 10 000 iterations, making their stochastic Galerkin model about 50 times faster. A direct comparison of computation time is not

attempted since Ge et al. (2008) used a second-order finite volume formulation that is inherently more expensive than the first-order formulation presented here.

Opportunities for Parallel Computation

Computation time for Monte Carlo and stochastic Galerkin simulations could be reduced by exploiting opportunities for parallelism. Monte Carlo simulations are called 'embarrassingly parallel' because, given sufficient processors, it is easy to perform iterations entirely in parallel. The stochastic Galerkin model is not embarrassingly parallel, but most operations can be parallelised nevertheless. Due to basis orthogonality, the evolution equation (equation 21) can be evaluated in parallel over basis functions $\Phi_0, \ldots, \Phi_P$. Gauss-Hermite quadrature of numerical fluxes (equation 22) can be parallelised, too. In this way, sampling the Riemann solver, which accounts for about 90% of the total computational cost, could be fully parallelised.

In theory, by fully exploiting parallelism, Monte Carlo and stochastic Galerkin simulations could be made to run almost as fast as a single iteration of the deterministic model, but this assumes access to sufficient processors. In practice, a fully parallelised Monte Carlo would require thousands of processors which are unavailable on typical hardware. In contrast, the stochastic Galerkin model with basis order $P = 3$ could be fully parallelised with just 16 processors using current, commodity hardware. With more hardware, additional parallelism could be achieved using domain decomposition techniques. The stochastic Galerkin method imposes no barriers to domain decomposition because it preserves the local element-wise operations of the underlying deterministic formulation. Given these substantial reductions in computation time and hardware demands, a stochastic Galerkin shallow flow model could alleviate the computational constraints associated with conventional Monte Carlo simulations (Neal et al., 2013), and allow probabilistic simulations to account for more sources of uncertainty.

Test 3: Steady-State Slow Flow Over an Irregular Bed

While the previous tests are restricted to idealised topography profiles, this final test simulates slow flow over a highly irregular bed that is more representative of real-world river hydraulics. The bed is defined with a range of uncertainty that characterises bathymetric measurement errors. Stochastic Galerkin model results are verified against the analytic solution.

For this purpose, following Tseng (2004), the 1D domain is $[0\,\text{m}, 1500\,\text{m}]$ with $M = 200$ elements such that the mesh spacing is $\Delta x = 7.5\,\text{m}$. The mean topography has an irregular profile as specified by Goutal and Maurel (1997) and is shown in figure 5a. The standard deviation of topography is $\sigma_z(x) = 0.5\,\text{m}$ across the entire domain, which is chosen since it provides an acceptable margin of error for local-scale flood modelling applications (Schumann and Bates, 2018). The initial free-surface elevation is 15 m with no uncertainty. Subcritical boundary conditions are imposed such that the mean upstream discharge per unit-width is $0.75\,\text{m}^2\,\text{s}^{-1}$ and the mean downstream water depth is 15 m, with no uncertainty on either boundary condition. Transmissive boundary conditions are used for the upstream water depth and downstream discharge. The test uses a timestep of $\Delta t = 0.5\,\text{s}$ and terminates at $t = 100\,000\,\text{s}$ when the water has converged on a steady state down to a convergence error of $10^{-8}\,\text{m}$ as measured by equation (44).

The stochastic Galerkin model is configured with a basis order $P = 3$, and the resulting steady-state solution is shown in figure 5. Since the flow is very slow, the free-surface elevation is uniform. The free-surface elevation has the same standard deviation as the topography, which is due to the water depth at the downstream boundary being fixed at 15 m with no uncertainty. As a result, the water depth profile is certain, and the uncertainty in the free-surface elevation is determined
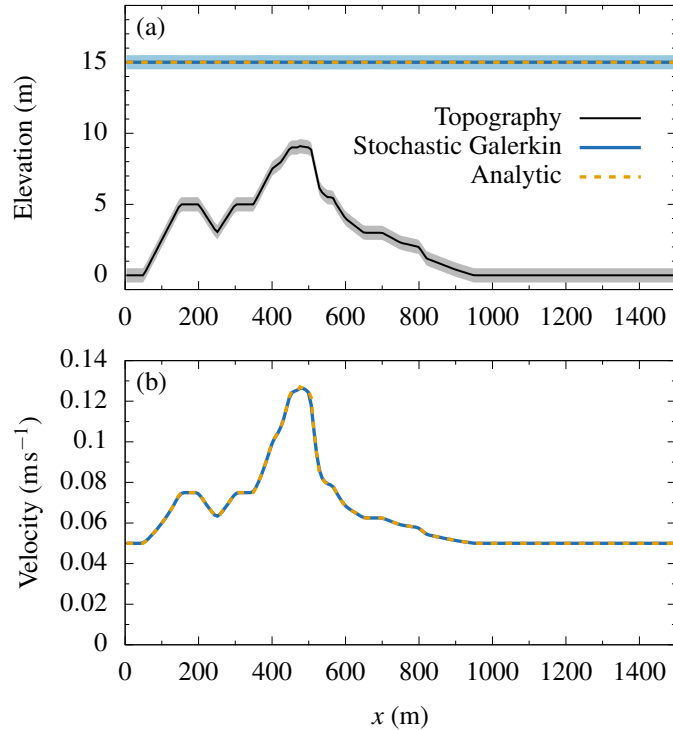
Fig. 5. Steady state solution over an irregular and uncertain bed. (a) Bed elevation and free-surface elevation. Mean values are marked with solid lines, with shading indicating plus or minus one standard deviation of uncertainty. (b) Spatial profile of depth-average velocity. The mean velocity is marked with a solid line and the standard deviation of velocity is negligible. In both panels, the analytic mean solution is marked with a dashed yellow line.

entirely by the topography. The corresponding depth-averaged velocity profile is shown in figure 5b. The stochastic Galerkin model calculates a velocity profile that is in excellent agreement with the analytic solution. The predicted velocity profile is certain with negligible standard deviation. This is expected since the water depth profile is unaffected by the prescribed topographic uncertainty. These results verify the robustness of the stochastic Galerkin model for flow over a highly-irregular, uncertain bed.

CONCLUSIONS AND OUTLOOK

This paper proposed a stochastic Galerkin shallow flow model using the surface gradient method to ensure well-balancing. The well-balancedness of the stochastic model was studied theoretically and verified in a series of idealised numerical tests, with a prescribed uncertainty that characterises topographic or bathymetric measurement errors. These tests demonstrated that the stochastic Galerkin model preserves discrete balance between flux and topography gradients to machine precision, and accurately reproduces steady flow over highly irregular and uncertain topography. Despite its relatively low-order Wiener-Hermite basis, the stochastic Galerkin model was able to simulate strongly nonlinear behaviour, producing discontinuous and highly non-Gaussian probability distributions that compared favourably against a Monte Carlo reference solution.

It is likely that a more sophisticated treatment using a multiresolution wavelet decomposition

Shaw, July 16, 2019

([Le Maître et al.](#), [2004](#); [Pettersson et al.](#), [2014](#)) or multi-element discretisation of the stochastic dimension ([Wan and Karniadakis](#), [2006](#); [Li and Stinis](#), [2015](#)) would further improve the representation of highly non-Gaussian and discontinuous probability distributions. It is also likely that a decomposition or discretisation of stochastic space would simplify a stochastic reformulation of positivity preserving schemes involving nonlinear operators, so enabling the development of stochastic wetting-and-drying processes that are essential for flood inundation modelling. Such improvements are the subject of future work.

The one-dimensional stochastic Galerkin model presented here is ideally suited to river routing applications: using as few as two expansion coefficients per element, the stochastic model can account for river bed measurement errors to produce a probabilistic flow simulation that is accurate and computationally efficient. Without parallelisation, the stochastic Galerkin model is at least 100 times faster than a Monte Carlo simulation. With parallelisation on commodity hardware, the computation time for the stochastic Galerkin model would be on par with a single iteration of the deterministic solver.

APPENDIX I. INSTRUCTIONS FOR RUNNING THE SHALLOW FLOW MODELS

Python 3 software is available for download from Zenodo (Shaw and Kesserwani, 2019b) which includes

1. a one-dimensional shallow flow model which can operate as a stochastic Galerkin model or a deterministic model
2. code for running Monte Carlo iterations of the deterministic model
3. preconfigured lake-at-rest, steady-state critical flow and steady-state flow over an irregular bed test cases

First ensure that Python 3, NumPy and SciPy are installed. Then, install the shallow flow model:

```
pip3 install --user --editable swepc.python
```

Running the Stochastic Galerkin Model

To run the stochastic Galerkin model:

```
swepc --degree <P> <testCase> <discretisation> --output-dir <directory>
```

where `<testCase>` is either `lakeAtRest`, `criticalSteadyState` or `tsengSteadyState`, `<discretisation>` is either `wellBalancedH` for the well-balanced surface gradient method, or `centredDifferenceH` for the centred difference method given by equations (42) and (43). If `--degree <P>` is omitted then the default basis order $P = 3$ is used. If `--degree 0` is specified then the model operates as a deterministic model. On running the model, the following text files are written to the specified output directory:

coefficients.dat Stochastic Galerkin expansion coefficients for topography z, water depth h and discharge q

statistics.dat Mean, standard deviation, skew and kurtosis calculated using equations (12), (13) and (14)

derived-statistics.dat Statistics for the depth-averaged velocity and the free-surface elevation, a derived quantity calculated from equation (18)

Each file has one line per element with explanatory header rows prefixed by `#`.

Calculating Probability Density Functions

A probability density function for any variable can be calculated from the expansion coefficients for a given element:

```
sed -n <line>p coefficients.dat
   | swepdf --min <value> --max <value> <variable> > pdf.dat
```

where `<line>` is the line number in `coefficients.dat` corresponding to the chosen element, and `<variable>` is `z` (topography), `water` (water depth), `q` (discharge) or `derived-eta` (free-surface elevation). The probability density function is calculated between the `--min` and `--max` values and written to `pdf.dat`.

Running Monte Carlo Simulations

To run a Monte Carlo simulation:

```
swepc --monte-carlo --mc-iterations <value>
  criticalSteadyState <discretisation> --output-dir <directory>
```

which writes the following text files:

statistics.dat Mean, standard deviation, skew and kurtosis for topography, water depth and discharge

derived-statistics.dat Mean and standard deviation for free-surface elevation and depth-averaged velocity

sample<n>.dat Deterministic model data with one file per element (numbered $0, \ldots, M-1$), each having one line per Monte Carlo iteration

REFERENCES

Audusse, E., Bouchut, F., Bristeau, M.-O., Klein, R., and Perthame, B. (2004). "A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows." *SIAM Journal on Scientific Computing*, 25(6), 2050–2065. doi:10.1137/S1064827503431090.

Bates, P. D., Pappenberger, F., and Romanowicz, R. J. (2014). "Uncertainty in flood inundation modelling." *Applied uncertainty analysis for flood risk management*, World Scientific, 232–269.

Bermúdez, A. and Vázquez, M. E. (1994). "Upwind methods for hyperbolic conservation laws with source terms." *Computers & Fluids*, 23(8), 1049–1071. doi:10.1016/0045-7930(94)90004-3.

Beven, K. and Binley, A. (1992). "The future of distributed models: model calibration and uncertainty prediction." *Hydrological processes*, 6(3), 279–298. doi:10.1002/hyp.3360060305.

Ge, L. and Cheung, K. F. (2011). "Spectral sampling method for uncertainty propagation in long-wave runup modeling." *Journal of Hydraulic Engineering*, 137(3), 277–288. doi:10.1061/(ASCE)HY.1943-7900.0000301.

Ge, L., Cheung, K. F., and Kobayashi, M. H. (2008). "Stochastic solution for uncertainty propagation in nonlinear shallow-water equations." *Journal of Hydraulic Engineering*, 134(12), 1732–1743. doi:10.1061/(ASCE)0733-9429(2008)134:12(1732).

Gerritsma, M., Van der Steen, J.-B., Vos, P., and Karniadakis, G. (2010). "Time-dependent generalized polynomial chaos." *Journal of Computational Physics*, 229(22), 8333–8363. doi:10.1016/j.jcp.2010.07.020.

Goutal, N. and Maurel, F. (1997). "Proceedings of the 2nd workshop on dam-break wave simulation." *Report No. HE-43/97/016/A*, Electricité de France, Département Laboratoire National d'Hydraulique, Groupe Hydraulique Fluviale.

Jin, S. (2001). "A steady-state capturing method for hyperbolic systems with geometrical source terms." *ESAIM: Mathematical Modelling and Numerical Analysis*, 35(4), 631–645. doi:10.1051/m2an:2001130.

Jin, S., Xiu, D., and Zhu, X. (2016). "A well-balanced stochastic Galerkin method for scalar hyperbolic balance laws with random inputs." *Journal of Scientific Computing*, 67(3), 1198–1218. doi:10.0.3.239/s10915-015-0124-2.

Jung, Y. and Merwade, V. (2012). "Uncertainty quantification in flood inundation mapping using generalized likelihood uncertainty estimate and sensitivity analysis." *Journal of Hydrologic Engineering*, 17(4), 507–520. doi:10.1061/(ASCE)HE.1943-5584.0000476.

Kim, B. and Sanders, B. F. (2016). "Dam-break flood model uncertainty assessment: case study of extreme flooding with multiple dam failures in Gangneung, South Korea." *Journal of Hydraulic Engineering*, 142(5), 05016002. doi:10.1061/(ASCE)HY.1943-7900.0001097.

Le Maître, O., Najm, H. N., Ghanem, R., and Knio, O. (2004). "Multi-resolution analysis of Wiener-type uncertainty propagation schemes." *Journal of Computational Physics*, 197(2), 502–531. doi:10.1016/j.jcp.2003.12.020.

Li, J. and Stinis, P. (2015). "Mesh refinement for uncertainty quantification through model reduction." *Journal of Computational Physics*, 280, 164–183. doi:10.1016/j.jcp.2014.09.021.

Liang, Q. and Marche, F. (2009). "Numerical resolution of well-balanced shallow water equations with complex source terms." *Advances in Water Resources*, 32(6), 873–884. doi:10.1016/j.advwatres.2009.02.010.

Montanari, A. and Di Baldassarre, G. (2013). "Data errors and hydrological modelling: The role of model structure to propagate observation uncertainty." *Advances in Water Resources*, 51, 498–504. doi:10.1016/j.advwatres.2012.09.007.

Neal, J., Keef, C., Bates, P., Beven, K., and Leedal, D. (2013). "Probabilistic flood risk mapping including spatial dependence." *Hydrological Processes*, 27(9), 1349–1363. doi:10.1002/hyp.9572.

Pettersson, P., Iaccarino, G., and Nordström, J. (2014). "A stochastic Galerkin method for the Euler equations with Roe variable transformation." *Journal of Computational Physics*, 257, 481–500. doi:10.1016/j.jcp.2013.10.011.

Roe, P. and Pike, J. (1984). "Efficient construction and utilisation of approximate riemann solutions." *Proc., Computing Methods in Applied Science and Engineering*, North-Holland Publishing Co., 499–518.

Sattar, A. M. and El-Beltagy, M. (2017). "Stochastic solution to the water hammer equations using polynomial chaos expansion with random boundary and initial conditions." *Journal of Hydraulic Engineering*, 143(2), 04016078. doi:10.1061/(ASCE)HY.1943-7900.0001227.

Schumann, G. J.-P. and Bates, P. D. (2018). "The need for a high-accuracy, open-access global DEM." *Frontiers in Earth Science*, 6, 225. doi:10.3389/feart.2018.00225.

Shaw, J. and Kesserwani, G. (2019a). "One-dimensional shallow flow solution data with uncertain topography." *Zenodo*, <https://doi.org/10.5281/zenodo.1469530> (Mar. 13, 2019).

Shaw, J. and Kesserwani, G. (2019b). "Python 3 implementation of a well-balanced one-dimensional stochastic galerkin shallow flow model with uncertain topography." *Zenodo*, <https://doi.org/10.5281/zenodo.1627619> (Mar. 13, 2019).

Toro, E. F. and Garcia-Navarro, P. (2007). "Godunov-type methods for free-surface shallow flows: A review." *Journal of Hydraulic Research*, 45(6), 736–751. doi:10.1080/00221686.2007.9521812.

Tryoen, J., Le Maître, O., Ndjinga, M., and Ern, A. (2010). "Intrusive Galerkin methods with upwinding for uncertain nonlinear hyperbolic systems." *Journal of Computational Physics*, 229(18), 6485–6511. doi:10.1016/j.jcp.2010.05.007.

Tseng, M.-H. (2004). "Improved treatment of source terms in TVD scheme for shallow water equations." *Advances in Water Resources*, 27(6), 617–629. doi:10.1016/j.advwatres.2004.02.023.

Wan, X. and Karniadakis, G. E. (2006). "Long-term behavior of polynomial chaos in stochastic flow simulations." *Computer Methods in Applied Mechanics and Engineering*, 195(41-43), 5582–5596. doi:10.1016/j.cma.2005.10.016.

Xiu, D. (2009). "Fast numerical methods for stochastic computations: a review." *Communications in Computational Physics*, 5(2-4), 242–272.

Xiu, D. and Karniadakis, G. E. (2002). "The Wiener–Askey polynomial chaos for stochastic differential equations." *SIAM Journal on Scientific Computing*, 24(2), 619–644. doi:10.1137/S1064827501387826.

Xiu, D. and Karniadakis, G. E. (2003). "Modeling uncertainty in flow simulations via generalized polynomial chaos." *Journal of Computational Physics*, 187(1), 137–167. doi:10.1016/S0021-9991(03)00092-5.

Zhou, J. G., Causon, D. M., Mingham, C. G., and Ingram, D. M. (2001). "The surface gradient method for the treatment of source terms in the shallow-water equations." *Journal of Computational Physics*, 168(1), 1–25. doi:10.1006/jcph.2000.6670.