

RESEARCH ARTICLE

Efficient Computation of Unsteady Flow in Complex River Systems with Uncertain Inputs

Nathan Gibson^{†*}, Christopher Gifford-Mears[‡], Arturo S. Leon[‡], and Veronika Vasylykivska[†]

[†]*Department of Mathematics, Oregon State University, 368 Kidder Hall, Corvallis, OR 97331-4605, USA;* [‡]*School of Civil and Construction Engineering, Oregon State University, 213 Owen Hall, Corvallis, OR 97331-3212, USA*

(January 30, 2013)

This paper examines the modeling and computational issues of a framework for representing uncertain inflows in river systems using the Polynomial Chaos approach. Ensemble forecasts are used to construct a Karhunen-Loève expansion of random inflows. The statistics of the stochastic outflow of the system are computed using Stochastic Collocation. The dynamics of the river system are efficiently simulated using the performance graphs approach.

Keywords:

Karhunen-Loève expansion, polynomial chaos, stochastic collocation, reservoir modeling

AMS Subject Classification:

65C20,35Q35,93E20

1. Introduction

Real-time operation of reservoir systems is important for many reasons, including water storage, electric power, flood control, recreation, water quality and downstream fishery needs. Uncertainties arise via upstream inflows, weather forecasts, imprecise measurements of water levels, and hydropower demands. The resulting PDE-constrained optimal control problem is a complex task involving stochastic inputs and objectives, probabilistic constraints, and nonlinear evolution equations imposed on massive domains. Both the optimization component and the uncertainty quantification require numerous forward simulations of the system. We focus here on the forward problem and limit the discussion to uncertain inputs.

Uncertain inflows were introduced into a multi-reservoir network using a linear, stochastic perturbation of an expected inflow hydrograph in [18]. A *polynomial chaos* (PC) expansion [5, 10, 29] of the outflow was computed using a *stochastic collocation* (SC) approach. The SC method can be interpreted to be a pseudospectral method which allows to approximate the multi-dimensional integrals in the stochastic space involved in a calculation of the coefficients of continuous orthogonal projection, using the Gaussian quadrature rule based on the collocation points chosen as the roots of suitable one-dimensional orthogonal polynomials [25, 27]. In the current work, we investigate additional aspects of this general approach to uncertainty quantification in reservoir modeling. In particular, we assume that

*Corresponding author. Email: gibsonn@math.oregonstate.edu

predictions of inflow hydrographs (ensemble forecasts) come from various sources, each of which possibly with its own probability of being realized. We wish to translate this discrete set of data into a continuous random framework amenable to PC expansion. Additionally, we wish to quantify the possible errors in the approximation resulting from this translation. We can then effectively reduce the dimension of the random input space to a manageable number with metrics to estimate the error induced by the approximate subspace. The use of the SC method in computing modes of the uncertain solution allows one to exploit the efficiencies in a deterministic forward simulation.

The approach for uncertainty quantification presented here will be included as part of an overall framework for solving the full uncertain constrained optimal control problem of reservoir planning and operations. Thus we must ensure that the methods are amenable to large scale optimization and parallelization. In particular, we mention that, in order to further reduce the computational burden of the uncertain forward problem, a complementary work in progress [12] involves developing an approach for decomposing the problem into subdomains (based on the work in [14]) which can be used in a parallelization of the deterministic forward simulation. Also, the PC expansion computed here will be used as an initial surrogate model for the evaluation of probabilistic constraints involving the components of the solution vector [13]. Combined with the methods which we describe below, the unified framework allows for efficient and adaptive determination of stochastic solutions to the uncertain multi-reservoir river system.

2. Governing Equations

In the following we present an unsteady flow routing (river system flow dynamics). Due to space limitations we consider only one-dimensional models. In a one-dimensional context, under a deterministic assumption, unsteady flows in open-channels are typically represented by the Saint-Venant equations, a pair of one-dimensional partial differential equations representing conservation of mass and momentum for a control volume, which is shown in conservative differential form in Equations (1) and (2)

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0, \quad (1)$$

$$\frac{1}{A} \frac{\partial Q}{\partial t} + \frac{1}{A} \frac{\partial}{\partial x} \left(\frac{Q^2}{A} \right) + g \cos(\theta) \frac{dy}{dx} - g(S_0 - S_f) = 0. \quad (2)$$

In these equations, x = distance along the channel in the longitudinal direction; t = time; Q = discharge; A = cross-sectional area; y = flow depth normal to x ; θ = angle between the longitudinal bed slope and a horizontal plane; g = acceleration of gravity; S_0 = bed slope and S_f = friction slope. Appropriate initial and boundary conditions are required to close the system. Due to the presence of non-linear terms in equation (2), there is, in general, no closed-form solution. The equations are therefore solved numerically. In a network involving numerous branches, the system of equations that must be solved becomes extremely large and the application of the full Saint-Venant equations becomes inefficient for real-time operation because of the significant computational requirements and error accumulations [16].

Instead we use the performance graphs approach described in [19]. The method solves a reduced non-linear system of equations to perform the hydraulic routing of

the system. The equations are assembled based on information in the reaches and nodes summarized in appropriate performance graphs formed from high fidelity, pre-computed solutions. These are combined with continuity and compatibility of water stages at junctions, and the system's initial and boundary conditions. Due to the pre-computation of solutions, efficiencies cannot be realized if the simulator must adjust to incorporate uncertainty, therefore we apply a *non-intrusive* uncertainty framework below.

3. Background

In the current work, only the stream inflows (external sources) are assumed to be completely stochastic. Other uncertain quantities are correlated to the uncertainty of the stream inflows using the dynamics of the system. For the efficient computation of the uncertainty components, rather than doing random sampling of the input distributions, we propose to explicitly model the random space (via random variables and processes) and perform a generalized Polynomial Chaos (PC) representation [9, 11, 27, 29] of the solutions.

3.1 Uncertainty Quantification Methods

Representation of the solutions in the form of a truncated PC expansion requires determining the coefficients of the expansion. One method for doing this is stochastic Galerkin (SG) method, see e.g., [27]. This approach results in a large coupled system of equations. The new system of equations must be discretized in space and time which means that the original deterministic solvers can not be used directly, since it is an *intrusive method* which changes the system to be solved. Instead, we wish to utilize a well-developed forward solution methodology based on performance graphs [19]. We therefore employ the SC method [25, 27] for the computation of coefficients of the PC expansion, a *non-intrusive method*, which we couple with the performance graphs implementation in OSU Rivers [19].

A popular alternative to SC is the classical Monte Carlo (MC) method. Both methods allow to utilize the readily available solvers corresponding to the deterministic equivalents of the system's governing equations (e.g., are *non-intrusive*). Each of the methods has its own advantages and disadvantages. In particular, while the MC methods are simple to implement, they require more simulations in general for the moments of the solution to converge. On the other hand, the convergence rate of relative L_2 errors of the mean and variance of the solutions obtained with PC expansions can be shown to be exponential as the number of basis functions increases [29]. The SC method is most appropriate for relatively small random dimension, which we take care to ensure with our methodology below.

Non-intrusive methods involving the construction of a polynomial approximation usually fall into one of the following three groups: interpolation, regression or pseudospectral projection. The main difference between these three approaches is that, in general, only the interpolation method requires the approximation to match the solution exactly at the collocation nodes, see [26, 27] for a detailed description of each of the approaches, together with the comparison of their advantages and weaknesses.

Interpolation is based on the construction of the Lagrange basis on a set of prescribed nodes in the random space. Although this approach is straightforward and easy to implement, the choice of the nodes is not trivial, especially in multidimensional spaces. Thus, it happens often that stochastic collocation methods based on

interpolation choose the nodes as a set of cubature points: full tensor [2, 28] or sparse grids [23].

Using the regression approach one estimates the PC coefficients by minimizing the mean square error of the response approximation. As in the case of interpolation, it also depends on the choice of the nodes. In [3, 24] authors investigate the use of the zeros of orthogonal polynomials as a starting point in the choice of the sampling nodes. They build a full tensor grid but choose only a prescribed number of nodes with the smallest norm. Sparse PC expansion based on regression is described in [4]. In the sparse PC expansion, fewer terms are kept in comparison with the full PC representation.

The pseudospectral approach is sometimes called a discrete projection method [27]. It is based on a numerical approximation of the coefficients of PC representation using quadrature formulas. This approach was first introduced in [25] where author also discusses different choices of the collocation nodes. A sparse alternative of the pseudospectral approach is presented in [7]. In particular, this paper addresses the error in the coefficients of PC expansion associated with the use of the sparse grid based quadrature rules and suggests a method which allows to minimize it. The pseudospectral approach looks for the coefficients associated with the known basis functions while interpolation approach looks for the basis functions corresponding to the known coefficients. Thus, from the point of view of implementation the pseudospectral approach is more appealing.

In our numerical experiments we use the pseudospectral approach to approximate the coefficients of the PC expansion. This method is straightforward and easy to implement. It allows the user control over the computational effort by allowing to compute only the coefficients which are important for a particular problem without evaluating the rest of the PC coefficients. Clearly, accuracy of the chosen quadrature rule is very important for this method.

The error associated with the use of quadrature approximated coefficients, rather than exact, in the PC representation is called an aliasing error, see e.g. [27]. In order to minimize this error high precision quadrature rules should be used. At the same time, in the case of high dimensionality of the random space, this would imply a significant increase of the computational work required. The use of sparse grids may decrease the computational work to a desired level. It is generally the case that when the number of random variables in the representation of the input parameters is greater than five, sparse grids outperform full tensor grids. Although we mention the possibility of using the sparse grids, a description of this approach is beyond the scope of this paper. As we show in our numerical results below, the size of the random space for the current problem does not exceed three, and furthermore, calculations based on full tensor grids do not require an unreasonable amount of time.

3.2 Related Efforts

PC methods have been studied in computational fluid dynamics by numerous investigators (e.g., [5, 15, 17, 29]). The *non-intrusive* SC method was introduced in the computational fluid dynamics literature in [21].

SC method was successfully applied to a non-linear model for incompressible flow and heat transfer around an array of circular cylinders based on the two-dimensional Reynolds-averaged Navier-Stokes equations [6]. The uncertainty was introduced through boundary conditions in a steady-state model.

A network of human arteries was considered in [30] where weakly non-linear 1D equations of pressure and flow wave propagation were used as a model in each

section of compliant vessels. While a network was considered, with mass balance interface conditions similar to the model described in the current work, the uncertain quantities were restricted to the geometric and physical properties of the artery, not inflows or boundary conditions. The study on human arteries did demonstrate the feasibility of SC on a physiologically realistic network of 37 branches.

4. Uncertainty Framework

The proposed framework can be used for any complex river network. For illustration purposes, consider the sample network system presented in Figure 1 from [18]. This dendritic-looped network consists of eight river reaches, two reservoirs and three boundary conditions (one inflow hydrograph, one stage hydrograph and one rating curve). The (nonlinear) relationship between variables

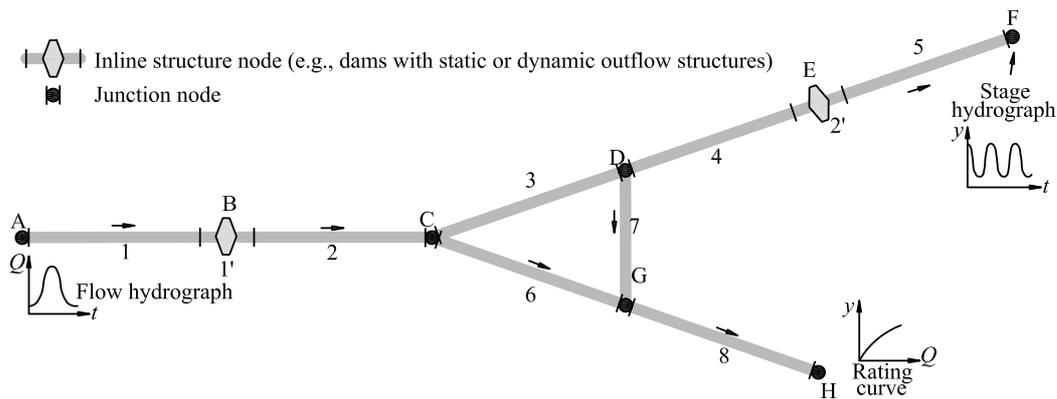


Figure 1. Schematic of a simple network system from [18]

$\vec{X} = [y_{d_1} \dots y_{d_s}, Q_{u_1} \dots Q_{u_s}, Q_{d_1} \dots Q_{d_s}]$ (water stages y and flow discharges Q upstream and downstream of each river reach) on each timestep is represented using the performance graphs approach described in [19].

In what follows we describe the method we use to introduce the uncertainty into the system. Let (Ω, \mathcal{F}, P) be a complete probability space, where Ω is the set of outcomes, $\mathcal{F} \subset 2^\Omega$ is a σ -algebra of events and $P : \mathcal{F} \rightarrow [0, 1]$ is a probability measure. Assume that the initial inflow function Q_{u_1} can be described as a function of finite number N_{rv} of independent random variables $\{\xi_k\}_{k=1}^{N_{rv}}$, i.e.

$$Q_{u_1}(t, \omega) = Q_{u_1}(t, \xi_1(\omega), \xi_2(\omega), \dots, \xi_{N_{rv}}(\omega)). \quad (3)$$

Let $\rho_k : \Gamma_k \rightarrow \mathbb{R}^+$, $k = 1, 2, \dots, N_{rv}$, denote the probability density function of the random variable ξ_k , with the image $\Gamma_k = \xi_k(\Omega) \subset \mathbb{R}$, $k = 1, 2, \dots, N_{rv}$. If the random variables $\{\xi_k\}_{k=1}^{N_{rv}}$ are independent then the joint probability density function ρ is given by the product of the corresponding densities

$$\rho(\mathbf{z}) = \prod_{k=1}^{N_{rv}} \rho_k(z_k), \quad \mathbf{z} \in \Gamma, \quad z_k \in \Gamma_k, \quad (4)$$

where $\Gamma = \prod_{k=1}^{N_{rv}} \Gamma_k \subset \mathbb{R}^{N_{rv}}$ is a support of the joint density function ρ . The introduction of uncertainty through the boundary conditions allows us to consider model (1) and (2) in the form of stochastic equations, i.e., find $Q : \mathbb{R} \times [0, T] \times \Gamma \rightarrow \mathbb{R}$

such that for all $\mathbf{z} \in \Gamma$, (1) and (2) hold subject to appropriate initial and boundary conditions, including $Q(x = 0, t, \omega) = Q_{u_1}(t, \omega)$.

5. Karhunen-Loève Representation of the Logarithm of the Inflow Function

In what follows we assume that the logarithm of the inflow function Q_{u_1} can be represented as a Gaussian process. This is quite a strong assumption, although the general uncertainty framework we use can be adjusted if it is violated.

In order to obtain a representation for the inflow function Q_{u_1} we use the following procedure [1]:

- (1) Suppose we have M realizations of the inflow function $\{Q_{u_1,i}\}_{i=1}^M$ measured at time points $\{t_j\}_{j=0}^n$, where $t_j = t_0 + jh$, $h = \frac{T - t_0}{n}$, $j = 1, \dots, n$, and $[t_0, T]$ is a time interval of interest. By $Q_{u_1,i}(t_j)$ we denote the value of the i -th realization of the inflow function at the time point t_j . Let $L_i(t_j) = \ln Q_{u_1,i}(t_j)$ denote the logarithm of the inflow at t_j , and $L(t) = \ln Q_{u_1,i}(t)$.
- (2) Then we compute the sample mean vector $\bar{L} = (\bar{L}_1, \bar{L}_2, \dots, \bar{L}_n)'$ and an $(n \times n)$ covariance matrix C with elements $c_{j,k}$ of the transformed inflows using the following formulas

$$\bar{L}_j = \bar{L}(t_j) = \frac{1}{M} \sum_{i=1}^M L_i(t_j), \quad c_{j,k} = \frac{1}{M-1} \sum_{i=1}^M (L_i(t_j) - \bar{L}_j)(L_i(t_k) - \bar{L}_k). \quad (5)$$

- (3) It follows that $L(t)$ can be represented in the form of its infinite series representation, called the Karhunen-Loève expansion [27],

$$L(t) = \bar{L}(t) + \sum_{k=1}^{\infty} \sqrt{\lambda_k} \psi_k(t) \xi_k, \quad (6)$$

where $\{\lambda_k, \psi_k\}_{k=1}^{\infty}$ are the eigenpairs of the integral equation

$$\lambda \psi(t) = \int_{t_0}^T C(s, t) \psi(s) ds, \quad (7)$$

with $C(t_j, t_k) = c_{j,k}$; and $\{\xi_k\}_{k=1}^{\infty}$ is a sequence of uncorrelated random variables with mean 0 and variance 1 defined by

$$\xi_k = \frac{1}{\sqrt{\lambda_k}} \int_{t_0}^T [L(t) - \bar{L}(t)] \psi_k(t) dt, \quad k \geq 1. \quad (8)$$

We assume that the eigenvalues are arranged in decreasing order, that is, $\lambda_1 > \lambda_2 > \lambda_3 > \dots$. In the case L is a Gaussian random process, $\{\xi_k\}_{k=1}^{\infty}$ are independent and identically distributed normal random variables with mean 0 and variance 1.

- (4) Then the inflow function Q_{u_1} has the following representation

$$Q_{u_1}(t) = \exp \left(\bar{L}(t) + \sum_{k=1}^{\infty} \sqrt{\lambda_k} \psi_k(t) \xi_k \right). \quad (9)$$

From the practical point of view it is not possible to use the infinite series representation of Q_{u_1} . The truncated representation is used instead

$$Q_{u_1}(t) \approx Q_{N_{rv}}(t) = \exp \left(\bar{L}(t) + \sum_{k=1}^{N_{rv}} \sqrt{\lambda_k} \psi_k(t) \xi_k \right). \quad (10)$$

The number of terms N_{rv} in the truncated representation can be chosen in different ways. One may use a fact that $\sum_{n=1}^{\infty} \lambda_n = \int_{t_0}^T C(s, s) ds$. Based on this criteria we can choose the number of terms that would capture the major part of the variability. Another way to determine N_{rv} is to look at the convergence rate of the eigenvalues and get rid of those that are close to 0, or insignificant in comparison with the first eigenvalue. For example, we can include the eigenvalues λ_n that satisfy

$$\lambda_n < a \lambda_1 \quad (11)$$

for some pre-defined constant $0 < a < 1$. In some sense a can be treated as a tolerance. A different perspective on this problem is given in the Section 7.

6. Polynomial Chaos Expansion

To solve the problem (1) and (2) in the stochastic context we form a generalized Polynomial Chaos (PC) expansion. To illustrate the idea of the proposed uncertainty approach, the following example is presented. Consider the quantity Q_{u_1} representing flow discharges upstream of reach 1 in Figure 1. We assume that based on the previous history or some additional data we can construct an uncertainty envelope around this prediction using the KL expansion of the logarithm $L = \ln Q_{u_1}$, for example,

$$Q_{N_{rv}}(t) = \exp \left(\bar{L}(t) + \sum_{k=1}^{N_{rv}} \sqrt{\lambda_k} \psi_k(t) \xi_k \right). \quad (12)$$

We want to determine the coefficients of a PC expansion of each component of the solution vector $[y_{d_1}, \dots, y_{d_s}, Q_{u_1}, \dots, Q_{u_s}, Q_{d_1}, \dots, Q_{d_s}]$, or some function of the solution vector. To do this, one may apply a pseudospectral approach in which one approximates the weighted inner products between the PC basis functions and a desired solution component with respect to the joint density ρ of the random variables in the representation of the inflow with a suitably chosen quadrature rule.

For example, consider the most downstream reach, Q_{d_s} . Its representation in terms of a degree p expansion

$$Q_{d_s}^P(t, \vec{\xi}) = \sum_{i=0}^{M_p} v_i(t) \Phi_i(\vec{\xi}), \quad (13)$$

where $\vec{\xi} = (\xi_1, \xi_2, \dots, \xi_{N_{rv}})$ is a vector of random variables in the representation of Q_{u_1} , $(M_p + 1)$ is a number of basis functions used. The functions $\{\Phi_i\}_{i=0}^{M_p}$ are the orthogonal polynomials of a degree at most p in each of N_{rv} variables. The maximum possible number of polynomial basis functions in this case is $p^{N_{rv}}$. Note that

the actual number of basis functions depends on if the same degree polynomials are used in each random dimension for the approximation. In our numerical experiments we use full tensor product basis. This implies that corresponding to the normal distribution of each components of $\vec{\xi}$, the orthogonal polynomials Φ_i can be chosen as the products of the corresponding univariate Hermite polynomials.

Since the relationship between Q_{u_1} and Q_{d_s} is clearly nonlinear, more than two basis functions will be required for accurate representation of Q_{d_s} . In the numerical simulations we construct the polynomial approximation of the solution of the degree $p = 2$ in each random variables. As we mentioned before depending on the importance of the particular component of the random vector $\vec{\xi}$ a different degree of approximation can be chosen in the k th random direction associated with ξ_k , $k = 1, \dots, N_{rv}$. This question is partially answered in the numerical experiments.

Each PC expansion coefficient can be found as an expectation

$$v_i(t) = E[Q_{d_s}(t, \vec{\xi})\Phi_i(\vec{\xi})] = \int_{\Gamma} Q_{d_s}(t, \mathbf{z})\Phi_i(\mathbf{z})\rho(\mathbf{z})d\mathbf{z}. \quad (14)$$

The computation of the coefficients (14) can be done efficiently with the use of the SC method [2].

The outline of the application of the SC method (pseudospectral approach) to the PC expansion is given below:

- (1) Choose a set of collocation points (\mathbf{z}_j, w_j) , $\mathbf{z}_j \in \Gamma$, where $\mathbf{z}_j = (z_{j,1}, z_{j,2}, \dots, z_{j,N_{cp}})$ is a j -th node and w_j is its corresponding weight, $j = 1, \dots, N_{cp}$. For the purpose of our numerical experiments we use the collocation points on the full tensor grid obtained as roots of univariate orthogonal polynomials with respect to the Gaussian density.
- (2) For each $j = 1, \dots, N_{cp}$ determine the inflow function $Q_{u_1,j}$ and solve the corresponding (deterministic) system of equations (1) and (2), in parallel, to obtain the flow $Q_{D_s,j}$.
- (3) Approximate the PC expansion coefficients

$$v_i(t) = \mathbb{E}[Q_{d_s}(t, \vec{\xi})\Phi_i(\vec{\xi})] \approx \sum_{j=1}^{N_{cp}} w_j Q_{d_s}(t, \mathbf{z}_j)\Phi_i(\mathbf{z}_j). \quad (15)$$

- (4) Finally, construct the N_{rv} -variate, p th-order PC approximation of the solution

$$Q_{d_s}^p(t, \vec{\xi}) = \sum_{i=0}^{M_p} v_i(t)\Phi_i(\vec{\xi}). \quad (16)$$

The same coefficients v_i can be used to approximate the first two moments of the solution, e.g.

$$E[Q_{d_s}(t, \vec{\xi})] \approx v_0(t), \quad \text{Var}[Q_{d_s}(t, \vec{\xi})] \approx \sum_{i=1}^{M_p} v_i(t)^2. \quad (17)$$

Gaussian quadrature applies efficiently to functions which can be represented as $g(\vec{\xi})W(\vec{\xi})$ where W is a weight function (e.g., the probability density function in an expected value) and $g(\vec{\xi})$ is well-approximated by a polynomial. The nodes $\vec{\xi}_j$ of the quadrature rule are the roots of a pre-determined orthogonal polynomial

(by choice of distribution) in the support of ρ , and the method has the highest degree of precision possible. SC method requires only solutions of the deterministic system evaluated at the fixed points $\{\vec{\xi}_j\}_{j=1}^{N_{cp}}$ of the random vector $\vec{\xi}$. Upon computation of the expansion coefficients for the quantities of interest, we have an analytical representation of a surrogate of the stochastic solution in polynomial form. This allows, among other things, various solution statistics to be easily obtained, such as expected value (or higher order modes), or parametric sensitivities [30]. The PC expansion for any function f of output (non-linear, non-smooth or even discontinuous) may be easily constructed as follows

$$v_i(t) = \mathbb{E}[f(\vec{X}(t, \vec{\xi}))\Phi_i(\vec{\xi})] \approx \sum_{j=1}^{N_{cp}} w_j f(\vec{X}(t, \vec{\xi}_j))\Phi_i(\vec{\xi}_j).$$

In practice, only desired functions of the solution of the system need to be represented explicitly. For instance, in a multi-objective optimal control framework, the outflows and water stages at the reservoirs may be the actual quantities of interest. The above example can be restated via a mapping from the solution quantities to the desired quantities. It is important to note that this mapping need not be linear, nor need it even be continuous, as demonstrated in [20], however in the latter case exponential convergence of the errors in the mean and variance of the solution with an increase in the number of basis functions is sacrificed in favor of algebraic.

7. Distributional Sensitivity

The representation of the random field in terms of the truncated series has its own features distinct from the original process. If the random process of interest is Gaussian, e.g., the logarithm of the inflow, then the truncated KL expansion is a random process represented as a linear combination of several standard Gaussian random variables. If the random process is not Gaussian, the representation of the process in the form of its Karhunen-Loève expansion becomes harder to obtain. The procedure has to involve the estimation of the distribution of the random coefficients in the series representation. In general, for non-Gaussian processes there may not be enough data to accurately specify the distribution of the random variables. This creates additional sources of uncertainty which in this case relate to the lack of data and may not be easily overcome.

In the work [22], the authors describe a distributional sensitivity analysis as a way to reduce epistemic uncertainty. The idea is to quantify the effect of the particular distribution of the random variables on the distribution or statistical moments of the solution. The random variables with large distributional sensitivity would require more attention and effort to approximate their distribution while the distribution of random variables with small distributional sensitivity can be approximated at lower computational expense.

The distributional sensitivity also allows one to obtain an additional rank (aside from the eigenvalues) of the random variables in the KL expansion based on their actual effect on the quantities of interest. In particular, relatively small values of the distributional sensitivity can suggest an insignificance of a particular random variable in the representation of the input, and consequently in the output. We illustrate this possibility in our numerical experiments. We note that other metrics exist for determining sensitivities in a PC expansion which also do not require additional model runs, c.f., [24]. Further, dimension adaptive sparse grids [8] could be used to determine which random dimensions to emphasize, however this approach

would not allow straight-forward parallelization of the quadrature nodes.

For simplicity of exposition we assume that the solution Q_{d_s} depends on the random vector $\vec{\xi} = (\xi_1, \xi_2, \dots, \xi_{N_{rv}})$ through the boundary conditions imposed as stream inflow Q_{u_1} . We assume that vector $\vec{\xi}$ has a joint density function ρ_1 . In our experiments ρ_1 is a joint Gaussian density, i.e. each component of the vector $\vec{\xi}$ has normal distribution with mean zero and variance 1. To quantify the sensitivity of the solution Q_{d_s} to the distribution of the random variables $\{\xi_k\}_{k=1}^{N_{rv}}$ we consider the following discrete distributional sensitivity

$$\text{DS}_{\mathcal{E}}[\rho_1, \rho_2](Q_{d_s}) = \frac{\|\mathcal{E}_{\rho_1}(Q_{d_s}) - \mathcal{E}_{\rho_2}(Q_{d_s})\|}{d(\rho_1, \rho_2)}, \quad (18)$$

where $\mathcal{E}_{\rho}(Q_{d_s})$ is a quantity of interest associated with Q_{d_s} , for example, mean or variance, with respect to the probability density ρ ; ρ_2 is a perturbation of the density ρ_1 ; $d(\rho_1, \rho_2)$ is a measure of distance between two densities, for example, it can be an L^1 norm.

It is worth mentioning that ρ_1 and ρ_2 do not necessarily share the common parameterization. The distributional sensitivity depends only on the densities ρ_1 and ρ_2 , so, in general, it does not matter what numerical methods are used to approximate the solution Q_{d_s} . The calculation of the distributional sensitivity is a post-processing step, no additional solutions are required. The moments can be obtained by using the SC method (described in the previous section). For the moments with respect to the density ρ_1 one can use the usual collocation points and weights; for the moments with respect to the perturbed density ρ_2 one can use the same collocation points with weights scaled by the ratio ρ_2/ρ_1 evaluated at the given collocation point. This means, for example, that if we approximate the expectation of Q_{d_s} with respect to the density ρ_1 with

$$E[Q_{d_s, \rho_1}](t) \approx \sum_{j=1}^{N_{cp}} w_j Q_{d_s}(t, \vec{\xi}_j), \quad (19)$$

then we can approximate the expectation of Q_{d_s} with respect to the density ρ_2 in the following way

$$E[Q_{d_s, \rho_2}](t) \approx \sum_{j=1}^{N_{cp}} w_j Q_{d_s}(t, \vec{\xi}_j) \rho_2(\vec{\xi}_j) / \rho_1(\vec{\xi}_j). \quad (20)$$

This approach allows to reuse the already available data and requires no additional simulations.

8. Computational Issues

As described above there are several aspects of the problem for which the computations can be quite expensive. In this simple model we have introduced only a single random inflow, while a realistic model of a complex river network might require several. Each inflow should be modeled with at least one random dimension. Taking into account all inflows would imply a dependence of the solution on a large number of variables. Computation of modes of a stochastic solution thus requires high dimensional integrations, in addition to the fact that each single deterministic simulation is already expensive.

In the above we have described two complementary approaches for reducing the computational effort required to obtain solutions to the uncertainty propagation problem. As the SC method described above uses pre-determined quadrature nodes, it is easily parallelizable into deterministic forward simulations. However, each forward problem using the performance graphs approach requires thousands of solutions to be stored in memory. Most massively parallel architectures are limited in memory and therefore not currently well-suited for this type of distributed computing as each computational node would need to hold an entire problem in memory. Therefore a fine-grained parallelism methodology, based on domain decomposition strategies [14], is being developed in a complementary effort [12]. We merely note here that the performance graph approach lends itself to this type of decomposition, and the non-intrusive nature of the SC method allows efficiencies in deterministic solvers to be maintained.

9. Numerical Experiments

For our simulation experiments we use the river system illustrated on the Figure 2. We assume that forecast of the inflow Q_{u_1} is given for reach 1. We wish to calculate the expected outflow $Q_{d_{25}}$ at reach 25, along with a quantification of uncertainty.

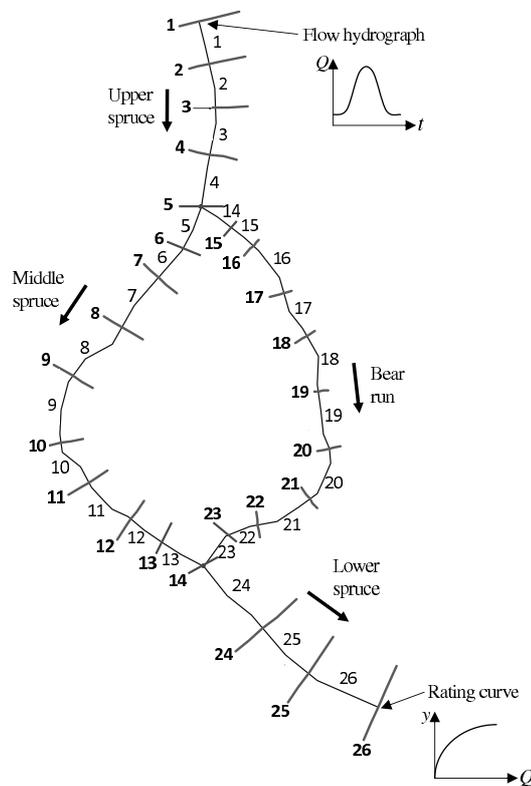


Figure 2. Schematic of a river system (from [19]) used in numerical experiments below

The predictions that we use are presented in Figure 3. We assume 10 ensembles (or predictions) of the stream inflow. This is meant to reflect the fact that in practice several competing forecasts are used to generate different scenarios.

We calculate the statistical mean and covariance of the data using equations (5). To find a spectral representation of the covariance function of the logarithm of the stream inflow based on its eigenvalues and eigenfunctions we solve the integral equation (7). The first five eigenvalues are presented in the Figure 4.(a). It is clear

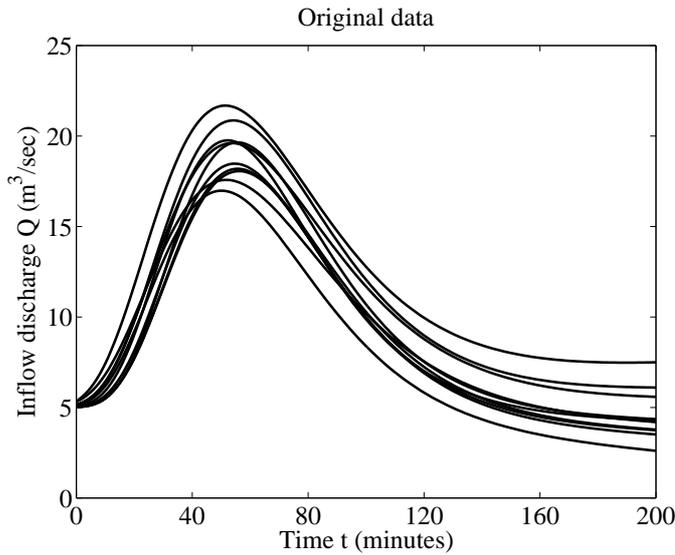


Figure 3. Original data

that only the first three eigenvalues are significant in terms of their magnitude: $\lambda_1 = 5.4721$, $\lambda_2 = 0.2658$, and $\lambda_3 = 0.0561$, $\lambda_4 = 0.0048$, $\lambda_5 = 9.849 \times 10^{-4}$. In particular, the first 3 eigenvalues contribute 99.8% of the variance of the infinite representation of the logarithm of the inflow function (recall $\sum_{k=1}^{\infty} \lambda_k = \int C(t, t) dt$). In the Figure 4.(c) we present only the eigenfunctions corresponding to the three largest eigenvalues.

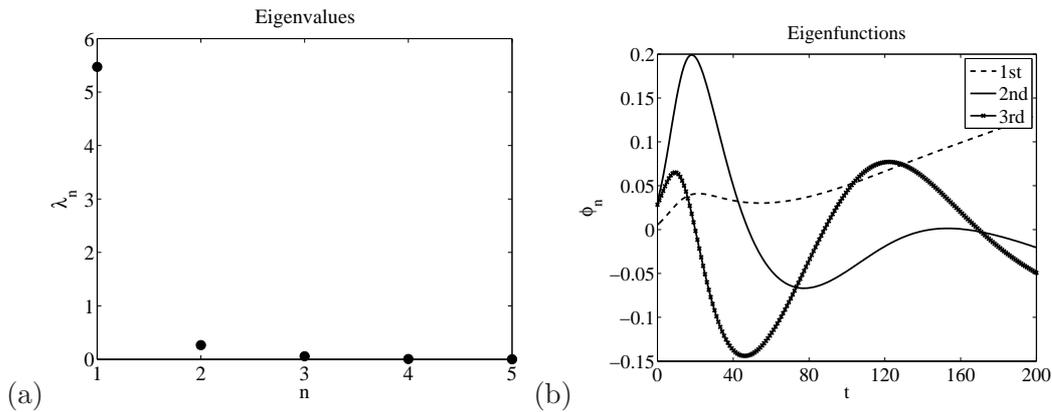


Figure 4. The eigenpairs obtained as part of the spectral representation of the data: (a) the five largest eigenvalues; (b) the first three eigenfunctions

Since only the first 3 eigenvalues are significant we use those to produce a truncated KL representation (10) of the logarithm of the stream inflow function $Q_{u_1}(t)$. We use the PC expansion coefficients to approximate the first two statistical moments of the outflow $Q_{d_{25}}$. For this demonstration, we employ 1-dimensional Gaussian quadrature points with five nodes in each of the three random dimensions (roots of the 5th degree univariate Hermite polynomial) to form a full tensor grid of 5^3 nodes. With this tensor grid we build the PC expansion of the second degree in each of the 3 random variables, e.g., with $3^3 = 27$ basis functions. Beyond evaluating the first two statistical moments of the solutions we lay the foundation to have an analytical expression of an approximation to the solution. This is necessary for the overall control framework in that it will be used as a surrogate model which allows to simplify the evaluation of the probabilistic constraints involving

the components of the solution vector [13]. These probabilistic constraints are outside the scope of the current paper. In Figure 5.(a) we present the realizations of outflow $Q_{d_{25}}$ evaluated at $5^3 = 125$ collocation points. We observe that they form five groups (and every group has five branches and 25 sub-branches arising due to the choice of the full tensor grid). In Figure 5.(b)-(c) we compare the mean and the standard deviation of the outflow when 1, 2 and 3 terms are included in the KL representation of the logarithm of the stream inflow. We observe the difference between the 1 and 2 terms representation but no visual difference between 2 and 3 terms representation. Note that for illustration purposes we have chosen an outflow as the solution, but, in general, similar analysis could be done for other components of the solutions.

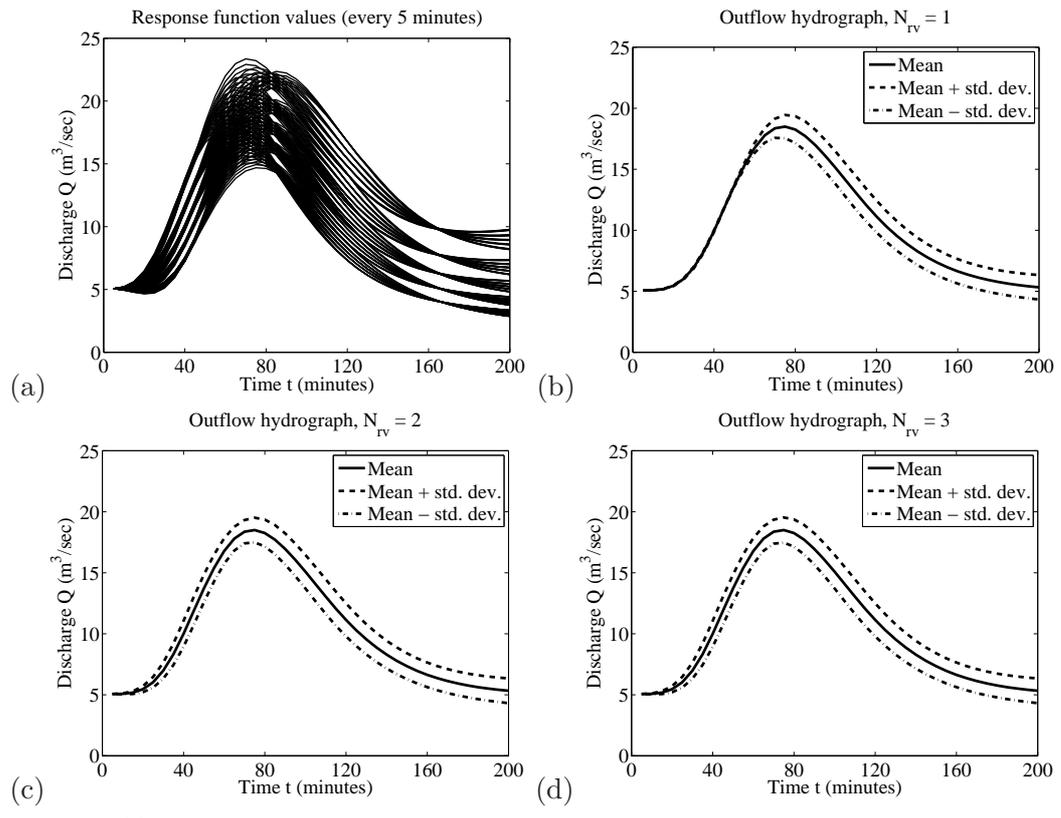


Figure 5. (a) The response outflow function evaluated at 125 collocation nodes with the interval of 5 minutes. Mean plus/minus standard deviation of the response function values: (b) only 1 term is included in the representation of the inflow; (c) 2 terms are included in the representation of the inflow; (d) 3 terms are included in the representation of the inflow

The magnitude of the eigenvalues shows the contribution of the corresponding term of the Karhunen-Loève expansion to the stream inflow. In other words it explains how much of the variation in data can be expressed with the particular term. The notion of the distributional sensitivity discussed earlier in this paper shows the contribution of each term to the outflow function. We measure the sensitivity (the effect of change of the distribution on the outflow) by perturbing the distribution of the random coefficients $\{\xi_k\}_{k=1}^3$ one at a time. For the perturbed version of the density ρ_1 we use the density ρ_2 corresponding to the normal distribution with mean δ and variance $1 + \varepsilon$.

We present our findings in two tables. To produce the results for Table 1 we assume that the KL representation of the logarithm of the stream inflow has two terms. The table shows the distributional sensitivity of the mean and variance of the outflow $Q_{d_{25}}$ due to the change of the mean and variance of each of the two

random variables. We observe that the mean of the solution $Q_{d_{25}}$ is affected more than the variance if only the means of the random variables are changed. The mean of the solution is only slightly sensitive to the variance of the random variables. Note that the sensitivity of the mean is well-approximated by perturbing both mean and variance simultaneously (the values are close in both cases), but the sensitivity for the variance is different in comparison with the first two cases. It appears in this case that the variance of the solution is more sensitive to the change in the distribution of the second variable. This suggests that the magnitudes of the sensitivity estimates are not necessarily consistent with the arrangement of the eigenvalues when both mean and variance of the random variables are changed but are consistent when only mean or variance is perturbed.

Table 1. Distributional sensitivity of expected outflow and its variance based on two terms in the KL expansion

ξ_i	δ	ϵ	$DS_{E[Q_{d_{25}}]}[\rho_1, \rho_2](Q_{d_{25}})$	$DS_{\text{Var}[Q_{d_{25}}]}[\rho_1, \rho_2](Q_{d_{25}})$
ξ_1	0.1	0	4.278	0.379
ξ_2	0.1	0	1.174	0.056
ξ_1	0	0.01	6.081e-3	2.068
ξ_2	0	0.01	7.913e-4	0.354
ξ_1	0.1	0.01	4.176	0.039
ξ_2	0.1	0.01	1.193	0.179

In Table 2 we consider a case when the representation of the logarithm of the stream inflow depends on three variables. The distributional sensitivity estimates for the first three variables seem to agree with the estimates presented in Table 2. The estimates for the third variable shows that they are at least 20 or more times smaller than the corresponding estimates for the second variable. These observations are consistent with the results presented in the Figure 5, that is, the statistical moments of the solution are not affected much by the presence and distribution of the third random variable.

Table 2. Distributional sensitivity of expected outflow and its variance based on three terms in the KL expansion

ξ_i	δ	ϵ	$DS_{E[Q_{d_{25}}]}[\rho_1, \rho_2](Q_{d_{25}})$	$DS_{\text{Var}[Q_{d_{25}}]}[\rho_1, \rho_2](Q_{d_{25}})$
ξ_1	0.1	0	4.281	0.381
ξ_2	0.1	0	1.174	0.058
ξ_3	0.1	0	0.047	0.001
ξ_1	0	0.01	6.085e-3	2.069
ξ_2	0	0.01	7.952e-4	0.355
ξ_3	0	0.01	9.331e-7	0.0006
ξ_1	0.1	0.01	4.176	0.039
ξ_2	0.1	0.01	1.193	0.181
ξ_3	0.1	0.01	0.047	0.0006

Note that the results in Tables 1 and 2 are obtained using the same set of data: no additional model runs are necessary to produce either the two or three terms case. All estimates were based on the original 125 model runs mentioned at the beginning of this section. The mean and variances of the solutions with respect to the perturbed distributions are calculated using the formula (20).

To compare to the moments of the solution $Q_{d_{25}}$ approximated with PC expansion, we calculate the moments obtained with 17000 MC realizations. We produce 17000 inflows and find the corresponding solutions $Q_{d_{25}}$. We also use varying numbers of collocation points to approximate the coefficients in the PC expansion. Each set of collocation points is used to construct a polynomial of the second degree in

each of the three random variables. The results are presented in the Figure 6. We see that PC expansions provide a good agreement with the moments obtained with MC method, which suggests that a second degree expansion gives an adequate representation of the solution $Q_{d_{25}}$.

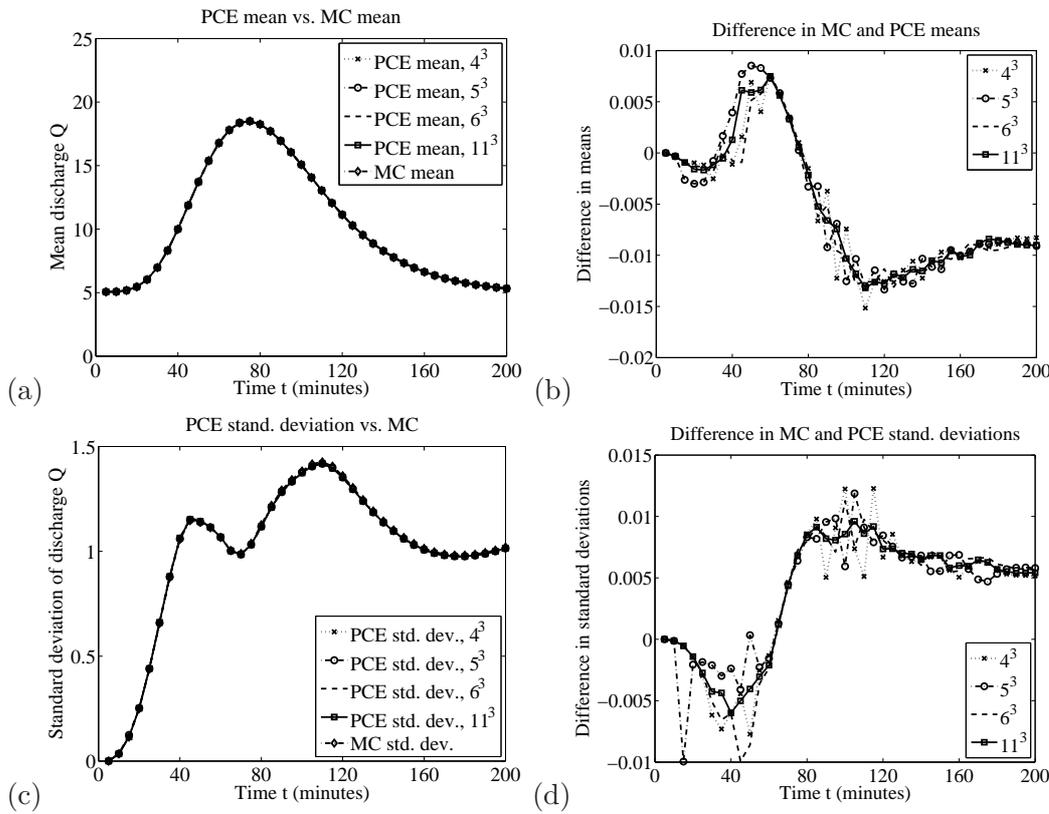


Figure 6. Comparison of the moments approximated with 17000 MC realizations and PC expansion with varying numbers of collocation points: (a) discharge means approximated by PC expansion and MC method; (b) difference in discharge means; (c) standard deviations approximated by PC expansion and MC method; (d) difference in the standard deviations of discharge.

In Figure 7.(a) we consider the magnitude of all 27 coefficients in the second degree full tensor PC approximation. The coefficients are calculated with the quadrature rule based on 11³ collocation points. We see that coefficient v_0 associated with a constant function $\Phi_0 = 1$ has the largest l_2 norm. The coefficients associated with the basis functions containing a linear term in either the first or second random variable are the next largest, and are an order of magnitude larger than the rest. For the illustration of convergence we approximate the first seven coefficients having the largest discrete l_2 norm with different number of collocation points (from 3³ to 7³) and compare those with the coefficients approximated with 11³ collocation points. The results are presented in Figure 7.(b). Indices shown in the legend represent the degree of the basis function in each of the random dimension, e.g., index $i = (i_1, i_2, i_3)$ denotes the basis function obtained as product of 3 functions: $\Phi_i = \Phi_{(i_1, i_2, i_3)} = \phi_{i_1}(\xi_1)\phi_{i_2}(\xi_2)\phi_{i_3}(\xi_3)$, where $\phi_0(\xi) = 1$, $\phi_1(\xi) = \xi$, and $\phi_2(\xi) = (\xi^2 - 1)/\sqrt{2}$. We assume that basis functions Φ_i are arranged in the graded lexicographic order, that is, index i is greater than index j if $\|i\|_1 \geq \|j\|_1$ and the first nonzero element in the difference, $i - j$, is positive. Note that the l_2 -error in the coefficients is already small using 3³ quadrature nodes, but it also exhibits convergence as the number of quadrature nodes is increased.

In Figure 8 we compare the moments (means and standard deviations) approximated with PC expansion coefficients based on 11³ collocation points to moments

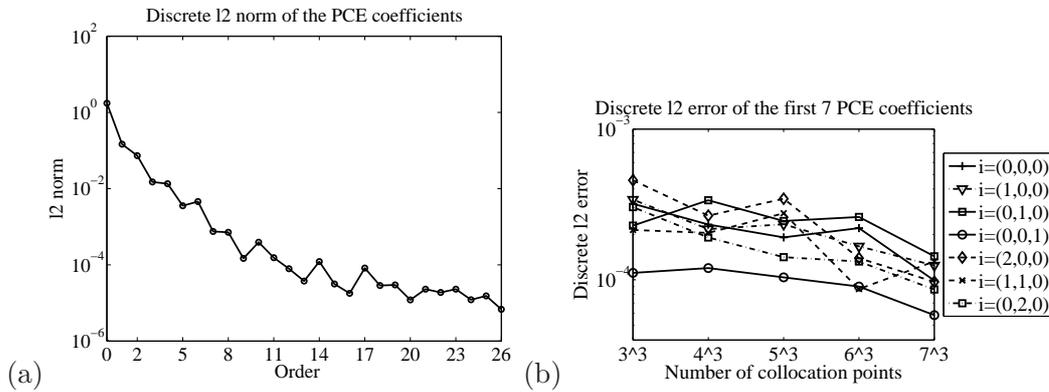


Figure 7. PCE coefficients: (a) discrete l_2 norm of all 27 coefficients (arranged in the graded lexicographic order) calculated with the quadrature rule based on 11^3 collocation points; (b) discrete l_2 error of the first seven PCE coefficients calculated with the quadrature rule based on varying numbers of collocation points, as compared to 11^3 .

approximated by the MC method (17000 realizations) and the PC approach using different numbers of collocation points. l_2 -convergence of the mean and standard deviation of the solution is demonstrated as the total number of function evaluations increased, but, as we see, the PC approach requires orders of magnitude fewer simulations than MC for the same level of accuracy.

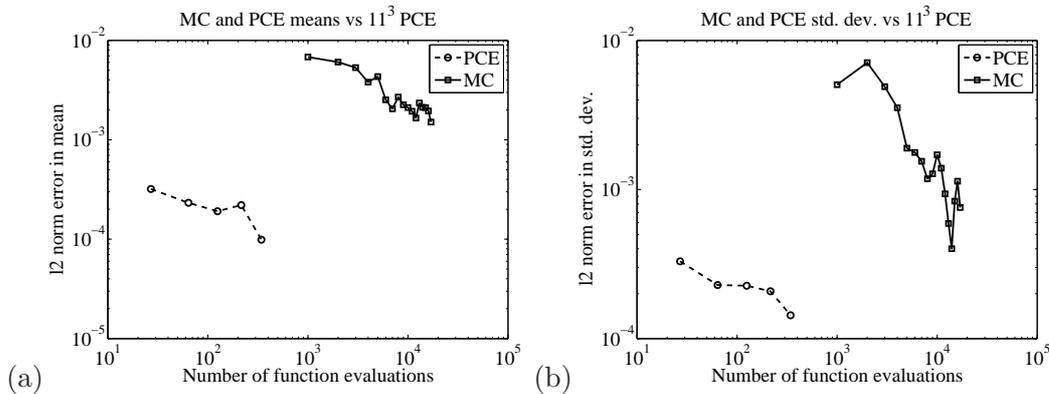


Figure 8. Discrete l_2 norm of the difference in moments approximated with MC method and PC based on different number of collocation points in comparison with PC moments based on 11^3 collocation points: (a) difference in means; (b) difference in standard deviations

10. Conclusions and Future Work

The work presented in this paper greatly extends the applicability of the research presented in [18]. We have used a Karhunen-Loève expansion-based representation of a space of random inflow functions implied by a given set of ensemble predictions. We also include the distributional sensitivity estimates to help quantify the importance of each random variable in the polynomial approximation of the outflow. In particular, distributional sensitivity suggests that 2 input random variables are sufficient. These results agree with the simulations. We also observed that quadratic polynomial representation of the output is adequate as moments approximated with MC and PC approach are close. We mentioned previously that the pseudospectral approach allows to evaluate the suitably chosen coefficients of the PC representation of a quantity of interest. Decay in the l_2 -norm of the coefficients has shown that only a few terms in the full tensor product are necessary. l_2 -error in the coefficients with the largest norms was shown to be sufficiently small using

3^3 quadrature nodes and illustrated a convergence as the number of quadrature nodes is increased. Lastly, l_2 -convergence of the mean and standard deviation of the solution obtained with PC method was demonstrated as the total number of simulations increased, requiring orders of magnitude fewer simulations than MC for the same level of accuracy.

Future work includes introducing inflow uncertainty into the full optimization framework, i.e., an optimal operation of multi-reservoir systems. Clearly, even a deterministic optimization problem of this complexity would require many forward simulations. With uncertainty included in the system, the computational effort increases dramatically. The uncertainty framework described in this work, together with a performance graph approach to unsteady flow routing, and fine-grained parallelism will be combined in order to attempt to reduce the computational expenses to practical levels.

A more complete introduction of uncertainty involves the stochastic representation of the price of electricity, load and wind power generation. Each of these sources of uncertainty have different structures and require additional theory to be developed.

Acknowledgments

This research was supported in part by the Bonneville Power Administration through the Technology Innovation Program, grant number TIP-258. The authors would like to thank an anonymous referee for many constructive suggestions.

References

- [1] I. Babuška, K.M. Liu, and R. Tempone, *Solving stochastic partial differential equations based on the experimental data*, Math. Models Methods Appl. Sci. 13 (2003), pp. 415–444, dedicated to Jim Douglas, Jr. on the occasion of his 75th birthday.
- [2] I. Babuška, F. Nobile, and R. Tempone, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM J. Numer. Anal. 45 (2007), pp. 1005–1034.
- [3] M. Berveiller, B. Sudret, and M. Lemaire, *Stochastic finite element: a non intrusive approach by regression*, Eur. J. Comput. Mech. 15 (2006), pp. 81–92.
- [4] G. Blatman and B. Sudret, *Adaptive sparse polynomial chaos expansion based on least angle regression*, J. Comput. Phys. 230 (2011), pp. 2345–2367.
- [5] Q. Chen, D. Gottlieb, and J. Hesthaven, *Uncertainty analysis for the steady-state flows in a dual throat nozzle*, J. Comput. Phys. 204 (2005), pp. 378–398.
- [6] P.G. Constantine, A. Doostan, and G. Iaccarino, *A hybrid collocation/Galerkin scheme for convective heat transfer problems with stochastic boundary conditions*, Internat. J. Numer. Methods Engrg. 80 (2009), pp. 868–880.
- [7] P.G. Constantine, M.S. Eldred, and E.T. Phipps, *Sparse pseudospectral approximation method*, Comput. Methods Appl. Mech. Engrg. 229 (2012), pp. 1–12.
- [8] T. Gerstner and M. Griebel, *Dimension-adaptive tensor-product quadrature*, Computing 71 (2003), pp. 65–87.
- [9] R.G. Ghanem and A. Doostan, *On the construction and analysis of stochastic models: characterization and propagation of the errors associated with limited data*, J. Comput. Phys. 217 (2006), pp. 63–81.
- [10] R.G. Ghanem and P.D. Spanos, *Stochastic finite elements: a spectral approach*, Springer-Verlag, New York (1991).
- [11] R.G. Ghanem, S.F. Masri, M.F. Pellissetti, and R. Wolfe, *Identification and prediction of stochastic dynamical systems in a polynomial chaos basis*, Comput. Methods Appl. Mech. Engrg. 194 (2005), pp. 1641–1654.
- [12] N.L. Gibson, L.A. Gomez, A.S. Leon, and V. Vasyukivska, *Domain decomposition approach for unsteady flow routing utilizing performance graphs*, To be submitted. 2013.
- [13] N.L. Gibson, M.G. McIntire, C. Hoyle, and V. Vasyukivska, *Robust reservoir operation with probabilistic constraints*, To be submitted. 2013.
- [14] M.D. Gunzburger, J.S. Peterson, and H. Kwon, *An optimization based domain decomposition method for partial differential equations*, Comput. Math. Appl. 37 (1999), pp. 77–93.
- [15] T.Y. Hou, W. Luo, B. Rozovskii, and H.M. Zhou, *Wiener chaos expansions and numerical solutions of randomly forced equations of fluid mechanics*, J. Comput. Phys. 216 (2006), pp. 687–706.
- [16] M.A. Hoy and A.R. Schmidt, *Unsteady Flow Routing in Sewers Using Hydraulic and Volumetric*

- Performance Graphs*, in *World Environmental and Water Resource Congress 2006: Examining the Confluence of Environmental and Water Concerns*, 2006.
- [17] O.M. Knio and O.P. Le Maître, *Uncertainty propagation in CFD using polynomial chaos decomposition*, *Fluid Dynam. Res.* 38 (2006), pp. 616–640.
- [18] A.S. Leon, N.L. Gibson, and C. Gifford-Miears, *Toward reduction of uncertainty in complex multi-reservoir river systems*, in *The XIX International Conference on Computational Methods in Water Resources*, 2012.
- [19] A.S. Leon, E.A. Kanashiro, and J.A. González-Castro, *A fast approach for unsteady flow routing in complex river networks based on performance graphs*, *J. Hydraul. Eng.* 139 (2013), pp. 284–295.
- [20] Y. Marzouk and D. Xiu, *A stochastic collocation approach to Bayesian inference in inverse problems*, *Commun. Comput. Phys.* 6 (2009), pp. 826–847.
- [21] L. Mathelin, M.Y. Hussaini, and T.A. Zang, *Stochastic approaches to uncertainty quantification in CFD simulations*, *Numer. Algorithms* 38 (2005), pp. 209–236.
- [22] A. Narayan and D. Xiu, *Distributional sensitivity for uncertainty quantification*, *Commun. Comput. Phys.* 10 (2011), pp. 140–160.
- [23] ———, *Stochastic collocation methods on unstructured grids in high dimensions via interpolation*, *SIAM J. Sci. Comput.* 34 (2012), pp. A1729–A1752.
- [24] B. Sudret, *Global sensitivity analysis using polynomial chaos expansions*, *Reliab. Eng. Syst. Saf.* 93 (2008), pp. 964–979.
- [25] D. Xiu, *Efficient collocational approach for parametric uncertainty analysis*, *Commun. Comput. Phys.* 2 (2007), pp. 293–309.
- [26] ———, *Fast numerical methods for stochastic computations: a review*, *Commun. Comput. Phys.* 5 (2009), pp. 242–272.
- [27] ———, *Numerical Methods for Stochastic Computations. A Spectral Method Approach*, Princeton University Press, Princeton, NJ (2010).
- [28] D. Xiu and J.S. Hesthaven, *High-order collocation methods for differential equations with random inputs*, *SIAM J. Sci. Comput.* 27 (2005), pp. 1118–1139.
- [29] D. Xiu and G.E. Karniadakis, *The Wiener-Askey polynomial chaos for stochastic differential equations*, *SIAM J. Sci. Comput.* 24 (2002), pp. 619–644 (electronic).
- [30] D. Xiu and S.J. Sherwin, *Parametric uncertainty analysis of pulse wave propagation in a model of a human arterial network*, *J. Comput. Phys.* 226 (2007), pp. 1385–1407.