Advances in Water Resources 111 (2018) 381–394

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Efficient uncertainty quantification in fully-integrated surface and subsurface hydrologic simulations

K.L. Miller[*,a], S.J. Berg[a,b], J.H. Davison[a,c], E.A. Sudicky[a,b], P.A. Forsyth[a,d]

[a] Aquanty Inc., Waterloo, ON, Canada
[b] Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, ON, Canada
[c] Department of Civil Engineering, Catholic University of America, Washington, DC, Canada
[d] Department of Computer Science, University of Waterloo, Waterloo, ON, Canada

ARTICLE INFO

Keywords:
HydroGeoSphere
Integrated hydrologic modeling
Polynomial chaos expansion
Uncertainty quantification

ABSTRACT

Although high performance computers and advanced numerical methods have made the application of fully-integrated surface and subsurface flow and transport models such as HydroGeoSphere common place, run times for large complex basin models can still be on the order of days to weeks, thus, limiting the usefulness of traditional workhorse algorithms for uncertainty quantification (UQ) such as Latin Hypercube simulation (LHS) or Monte Carlo simulation (MCS), which generally require thousands of simulations to achieve an acceptable level of accuracy. In this paper we investigate non-intrusive polynomial chaos for uncertainty quantification, which in contrast to random sampling methods (e.g., LHS and MCS), represents a model response of interest as a weighted sum of polynomials over the random inputs. Once a chaos expansion has been constructed, approximating the mean, covariance, probability density function, cumulative distribution function, and other common statistics as well as local and global sensitivity measures is straightforward and computationally inexpensive, thus making PCE an attractive UQ method for hydrologic models with long run times. Our polynomial chaos implementation was validated through comparison with analytical solutions as well as solutions obtained via LHS for simple numerical problems. It was then used to quantify parametric uncertainty in a series of numerical problems with increasing complexity, including a two-dimensional fully-saturated, steady flow and transient transport problem with six uncertain parameters and one quantity of interest; a one-dimensional variably-saturated column test involving transient flow and transport, four uncertain parameters, and two quantities of interest at 101 spatial locations and five different times each (1010 total); and a three-dimensional fully-integrated surface and subsurface flow and transport problem for a small test catchment involving seven uncertain parameters and three quantities of interest at 241 different times each. Numerical experiments show that polynomial chaos is an effective and robust method for quantifying uncertainty in fully-integrated hydrologic simulations, which provides a rich set of features and is computationally efficient. Our approach has the potential for significant speedup over existing sampling based methods when the number of uncertain model parameters is modest (≤ 20). To our knowledge, this is the first implementation of the algorithm in a comprehensive, fully-integrated, physically-based three-dimensional hydrosystem model.

1. Introduction

As a result of today's high performance computers, sophisticated mathematical models are capable of incorporating many complex processes. As is often the case, the partial or incomplete knowledge of these processes and the input parameters required to describe them necessitates the analyst to make various assumptions and approximations, and in doing so, introduces uncertainty into the model. The aim of uncertainty quantification is to estimate the variability in model responses propagated by model uncertainty. Consequently, uncertainty quantification provides the modeler with a certain level of confidence regarding the predictions made by their model. Among the many different types of uncertainty present in a mathematical model, we consider *parametric uncertainty*, that is, the uncertainty in model responses that arises from uncertainty in model input parameters.

In practice, Monte Carlo simulation (MCS) (Metropolis and Ulam, 1949) or Latin Hypercube simulation (LHS) (Iman et al., 1981a; 1981b) are the workhorse algorithms for uncertainty quantification.

* Corresponding author.
E-mail addresses: kmiller@aquanty.com (K.L. Miller), sberg@aquanty.com (S.J. Berg), jdavison@aquanty.com (J.H. Davison), esudicky@aquanty.com (E.A. Sudicky), paforsyt@uwaterloo.ca (P.A. Forsyth).

https://doi.org/10.1016/j.advwatres.2017.10.023
Received 13 June 2017; Received in revised form 13 October 2017; Accepted 17 October 2017
Available online 18 October 2017

These algorithms are robust and scale well to higher dimensions; however, they are slow to converge, generally requiring thousands of simulations to obtain an acceptable level of accuracy. In contrast to random sampling methods, a polynomial chaos expansion represents a model response of interest as a weighted sum of polynomials over the random inputs. A non-intrusive formulation treats the underlying deterministic model as a black box and consequently does not require any modification of the deterministic model. Depending on the nature of the problem at hand, polynomial chaos has the potential for significant speedup over more traditional random sampling methods. Moreover, once an expansion has been constructed, approximating the mean, covariance, probability density function, cumulative distribution function, and other common statistics as well as local and global sensitivity measures is straightforward and computationally inexpensive.

Polynomial chaos has its roots in the *homogeneous chaos* introduced by Wiener (1938), which uses Hermite polynomials to model random processes involving normal random variables. Ghanem and Spanos (1991) pioneered the very successful *stochastic finite elements*, which combines a Hermite polynomial basis with a finite element approach as a method to analyze uncertainty in solid mechanics. Their method was later generalized by Xiu and Karniadakis (2002) to the *generalized polynomial chaos*, based on the correspondence between certain probability density functions and weight functions corresponding to orthogonal polynomials from the Askey scheme (Askey and Wilson, 1985). It should be noted that a similar method referred to as *probabilistic collocation* was proposed much earlier by Tatang et al. (1997) and Isukapalli et al. (1998); however, their approach defines the expansion coefficients through the solution of a least-squares regression problem instead of via pseudo-spectral projection onto an orthogonal polynomial basis. A related method that employs a basis of Lagrange interpolating polynomials referred to as *stochastic collocation* was proposed in Mathelin and Hussaini (2003) and has been further developed in Xiu and Hesthaven (2005), Nobile et al. (2008a); 2008b) and Babuška et al. (2007). Generalizations and extensions of these methods have been proposed in the literature (Blatman and Sudret, 2010; Foo et al., 2008; Wan and Karniadakis, 2005; 2006).

In this paper, we investigate non-intrusive polynomial chaos for quantification of uncertainty in fully-integrated, physics-based hydrologic simulations using the HydroGeoSphere (HGS) model (Aquanty Inc., 2015). As part of our implementation details, we discuss an iterative refinement technique that incrementally improves a chaos expansion through optimal reuse of previous lower-order expansions. The effectiveness of polynomial chaos as a tool for uncertainty quantification has been demonstrated in many areas including computational fluid dynamics (Hosder et al., 2007; 2006), single and multiphase flow in heterogeneous media (Li and Zhang, 2007; 2009), vehicle dynamics (Kewlani et al., 2012), groundwater modeling (Deman et al., 2016), loosely coupled surface/subsurface modeling (Wu et al., 2014), and seawater intrusion modeling (Rajabi et al., 2015; Riva et al., 2015). However, to the best of our knowledge, this is the first time that it has been used to quantify the uncertainty in simulations generated by a globally implicit, fully-integrated surface and variably-saturated subsurface flow and solute transport model. Our polynomial chaos implementation was validated by comparison with analytical solutions as well as solutions obtained via LHS for simple numerical problems. It was then used to quantify the parametric uncertainty in a series of numerical problems with increasing complexity, including: a two-dimensional (2D) fully-saturated, steady-state flow and transient transport problem with six uncertain parameters and one quantity of interest (Section 7.1); a one-dimensional (1D) variably-saturated column test involving transient flow and transport, with four uncertain parameters, and two quantities of interest at 101 spatial locations and five different times each (1010 total) (Section 7.2); and a three-dimensional (3D) fully-integrated surface and subsurface flow and transport problem for a small catchment with seven uncertain parameters and three quantities of interest at 241 different times each (Section 7.3).

The main contributions of this manuscript include:

- The development of a model independent, robust, and user friendly uncertainty quantification code based on polynomial chaos for application to hydrologic simulations. In particular, we present a novel iterative refinement procedure for incrementally improving regression-based polynomial chaos expansions through optimal reuse of previous lower-order expansions.
- The novel application of our code to a highly nonlinear fully-integrated surface water/groundwater problem. In particular, we demonstrate that PCE can be applied to this complex model problem to efficiently compute time-varying global sensitivity indices that provide insight into the model behavior as well as the physical system as a whole.

The remainder of this paper is organized as follows: Sections 2 and 3 introduce the notation and setup the mathematical framework. In Section 4, we provide a thorough overview of polynomial chaos including a brief discussion of the computation of statistics and sensitivity indices from an expansion. Section 5 discusses the details of the implementation and in particular highlights the iterative refinement procedure. Section 6 describes the HydroGeoSphere model, and Section 7 describes the numerical testing and provides a discussion of the results. Concluding statements are presented in Section 8.

2. Notation

Throughout this paper we adhere to the following notational conventions. Uppercase and lowercase letters denote scalar quantities and in some cases may be used to denote sets, functions, or operators. Bold lowercase letters always denote vectors (e.g., $\mathbf{x}$) and bold uppercase letters always denote matrices (e.g., $\mathbf{A}$). Uppercase script letters always denote sets or spaces (e.g., $\mathscr{A}$, $\mathscr{D}$). The set of nonnegative integers $\{0, 1, 2, \ldots\}$ is denoted by $\mathbb{N}$, the set of strictly positive integers $\{1, 2, 3, \ldots\}$ is denoted by $\mathbb{N}_1$, and the set of real numbers is denoted by $\mathbb{R}$. The set of all n-vectors with elements in some set $\mathscr{B}$ for some positive integer n is denoted by $\mathscr{B}^n$. We use $U(a, b)$ to denote the uniform probability distribution on the interval $[a, b]$, $\mathrm{LogN}(\mu, \sigma)$ to denote the lognormal distribution with location parameter μ and scale parameter σ, and $\mathrm{LogU}(a, b)$ to denote the loguniform distribution on the interval $[a, b]$. We note that a random variable $X \sim \mathrm{LogU}(a, b)$ if and only if $\ln X \sim U(\ln a, \ln b)$.

3. Mathematical framework

Consider a mathematical model that depends on a finite collection of parameters $\Xi = \{\xi_1, \ldots, \xi_d\}$ and let $u(\mathbf{x}, t)$ be any real-valued response of this model, where $\mathbf{x} \in \mathbb{R}^n$ is the position variable and $t \geq 0$ is the time variable. Suppose that the model parameters Ξ are uncertain and that we would like to quantify the resulting uncertainty in u. We refer to Ξ as the *parameters of interest* and to u as a *quantity of interest*. Note that u may also depend on additional "certain parameters" that are fixed and are not considered by our analysis. A natural way to approach the mathematical formulation of this problem is to adopt a probabilistic framework for the uncertain parameters, treating them as random variables, and recasting the deterministic function u as a function of these random variables. In doing so, the uncertainty in u may then be rigorously quantified through statistical measures such as its mean and variance.

The random variables $\xi_1, \ldots, \xi_d$ are modeled as a d-variate random vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_d)$ in a properly defined probability space $(\Omega, \mathscr{F}, P)$, where Ω is the sample space, $\mathscr{F}$ is the event space, P is the probability measure, and $\boldsymbol{\xi}: \Omega \to \mathbb{R}^d$. We make the assumption that $\boldsymbol{\xi}$ has independent components. We note that as discussed in Eldred and Burkardt (2009), this assumption is not absolutely necessary; in theory,