Journal of Hydrology 529 (2015) 1129–1146

Contents lists available at ScienceDirect

Journal of Hydrology

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

A fractional factorial probabilistic collocation method for uncertainty propagation of hydrologic model parameters in a reduced dimensional space

S. Wang [a], G.H. Huang [a,*], W. Huang [b], Y.R. Fan [a], Z. Li [a]

[a] Faculty of Engineering and Applied Science, University of Regina, Regina, Saskatchewan S4S 0A2, Canada
[b] Department of Civil Engineering, McMaster University, Hamilton, ON L8S 4L7, Canada

ARTICLE INFO

Article history:
Received 30 July 2014
Received in revised form 10 September 2015
Accepted 14 September 2015
Available online 25 September 2015
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Vazken, Associate Editor

Keywords:
Probabilistic collocation
Hydrologic model parameters
Interaction
Uncertainty propagation
Polynomial chaos expansion
Multi-level factorial design

SUMMARY

In this study, a fractional factorial probabilistic collocation method is proposed to reveal statistical significance of hydrologic model parameters and their multi-level interactions affecting model outputs, facilitating uncertainty propagation in a reduced dimensional space. The proposed methodology is applied to the Xiangxi River watershed in China to demonstrate its validity and applicability, as well as its capability of revealing complex and dynamic parameter interactions. A set of reduced polynomial chaos expansions (PCEs) only with statistically significant terms can be obtained based on the results of factorial analysis of variance (ANOVA), achieving a reduction of uncertainty in hydrologic predictions. The predictive performance of reduced PCEs is verified by comparing against standard PCEs and the Monte Carlo with Latin hypercube sampling (MC-LHS) method in terms of reliability, sharpness, and Nash–Sutcliffe efficiency (NSE). Results reveal that the reduced PCEs are able to capture hydrologic behaviors of the Xiangxi River watershed, and they are efficient functional representations for propagating uncertainties in hydrologic predictions.

1. Introduction

Hydrologic models make use of simple mathematical equations to conceptualize physical behaviors of natural systems that involve complex interactions between the atmospheric, land surface, and subsurface components of the water cycle (Moradkhani et al., 2005). Due to the spatial heterogeneity of hydrologic systems and the scarcity of data, many parameters that represent hydrologic properties cannot be exactly identified. Moreover, certain parameters may be well-defined at a point-scale but not a mesh-scale or at a catchment scale, resulting in poor parameter identifiability. Thus, hydrologic parameters are often modelled as random variables. When uncertainties exist in the form of random variables, effective characterization and propagation are crucial for robust hydrological modelling (Konda et al., 2010).

Uncertainty propagation in various stochastic systems has been extensively studied over the past decade (Kunstmann and Kastens, 2006; Le Maître et al., 2004; Oladyshkin et al., 2011; Quintero et al., 2012; Wang et al., 2012, 2013b; Yen et al., 2014; Zhang and Sahinidis, 2012). Monte Carlo methods and their variants (e.g., Latin hypercube sampling) are some of the most widely used methods for uncertainty analysis (Oladyshkin and Nowak, 2012). These methods rely on repeated random sampling to obtain numerical results expressed as probability distributions. Thus their accuracy depends on the number of realizations of random parameters. The brute-force Monte Carlo simulation is straightforward to implement; however, its main limitation is the requirement of a large computational power, especially for large-scale problems (He et al., 2012; Li and Zhang, 2007; Wang et al., 2014).

As an attractive alternative, polynomial chaos expansion (PCE) techniques that originate from the homogeneous chaos theory proposed by Wiener (1938) represent a stochastic process by a spectral expansion based on Hermite orthogonal polynomials in terms of Gaussian random variables (Lin and Tartakovsky, 2009). For stochastic systems that involve physical parameters with non-Gaussian distributions, Xiu and Karniadakis (2002) further developed a generalized PCE to represent various stochastic processes with different types of orthogonal polynomials in order to

* Corresponding author. Tel.: +1 306 585 4095; fax: +1 306 585 4855.
E-mail addresses: sshuo.wwang@gmail.com (S. Wang), huangg@uregina.ca (G.H. Huang).

http://dx.doi.org/10.1016/j.jhydrol.2015.09.034
0022-1694/

achieve the optimal convergence rate. PCE offers an efficient way of characterizing nonlinear effects in stochastic analysis and represents the dependency of model outputs on uncertain input parameters by a set of high-dimensional orthogonal polynomials. This is achieved by projecting the model response surface onto a basis of orthogonal polynomials in the probabilistic parameter space (Oladyshkin et al., 2011). PCE-based methods have been extensively used in various fields, such as transport and flow in porous media (Ghanem, 1998; Laloy et al., 2013; Liao and Zhang, 2013), computational fluid dynamics (Mathelin et al., 2005; Najm, 2009), multibody dynamic systems (Sandu et al., 2006), as well as environmental and biological systems (Isukapalli et al., 1998).

In terms of the propagation of uncertainty from model parameters to outputs, PCE falls into two categories for the involved projection integrals: intrusive and non-intrusive approaches (Najm, 2009). The intrusive approaches rely on a Galerkin-projection reformulation of the original model equations to solve for the coefficients in the expansion (Ghanem and Spanos, 1991). These approaches require reformulating and solving a coupled system of deterministic and ordinary differential equations, which may become complex and cumbersome due to necessary symbolic manipulations, especially when there are a large number of the PCE coefficients to be determined. The non-intrusive approaches evaluate the coefficients in the expansion by employing deterministic sampling of the original model that can be treated as a black box, and require no manipulation of underlying partial differential equations. Thus, the non-intrusive approaches have been receiving increasing attention. Tatang et al. (1997) developed a probabilistic collocation method (PCM), which was particularly useful for uncertainty analysis of the computationally demanding problems because the PCM was nonintrusive and its implementation was relatively straightforward. The PCM can be used to compute the PCE coefficients by using the model outputs at selected collocation points. However, since the number of required collocation points is always smaller than the number of available Gaussian quadrature points, it is difficult to select the appropriate collocation points to construct the functional approximation (Wei et al., 2008). To address the shortcoming of the standard collocation technique, Isukapalli et al. (1998) proposed a stochastic response surface method to evaluate the PCE using a regression-based method. The regression method facilitates better approximations of model outputs by performing a larger number of model simulations. However, the number of sample points used in the regression method must be twice the number of unknown coefficients in order to obtain robust estimates, resulting in an enormous computational demand (Isukapalli, 1999).

In recent years, the PCM and its variants have been successfully applied to various fields (Fajraoui et al., 2011; Li et al., 2009; Müller et al., 2011; Zheng et al., 2011). For example, Shi et al. (2009) took advantage of the PCM to study the nonlinear flow in heterogeneous unconfined aquifers. Liao and Zhang (2013) proposed a new location-based transformed PCM for the stochastic analysis of geophysical models under strongly nonlinear conditions. Sun et al. (2013) used the PCM to assess leakage detectability at geologic CO_2 sequestration sites under parameter uncertainty. In previous studies, the existing methods often suffer from issues resulting from high dimensionality. As the number of random variables increases, the number of available collocation points increases exponentially. The PCM then becomes unstable for evaluating the PCE with a high-dimensional parameter space. This is because the response surface has to intersect all the collocation points in order to well represent the relationship between model inputs and outputs in terms of random variables, and each of the collocation points in the model space can significantly affect the behavior of the orthogonal polynomial. Further, to ensure the robustness of the PCM, all combinations of sample points chosen according to the Gaussian quadrature rule should be taken into account for estimating the coefficients of the PCE. However, such a collocation scheme may become too complicated and computationally unrealistic.

To address the above issues, the objective of this study is to propose a computationally efficient fractional factorial probabilistic collocation method for propagation of uncertainties in hydrologic model parameters. Since each term in the PCE has different contributions to the variability of the functional response, the proposed method is capable of revealing statistical significance of linear, nonlinear (i.e., second- and higher-order), and interaction PCE terms. The terms that have little effects on the outputs of the functional approximation can thus be discarded, leading to a reduced PCE only with statistically significant terms. Such a truncated expansion in orthogonal polynomials is necessary to ease the computational effort, especially for large-scale stochastic hydrologic systems with a high-dimensional parameter space. The proposed methodology will be applied to the Xiangxi River watershed by using the HYMOD conceptual hydrologic model to demonstrate its validity and applicability, as well as its capability of revealing mechanisms embedded within a number of hydrological complexities. A set of reduced PCEs will be obtained for streamflow simulations, and the derived uncertainty intervals of daily streamflows will then be compared against those from standard PCEs as well as those from the MC-LHS method. The probabilistic performance measures of reliability and sharpness will be used to evaluate the predictive capacities of different approaches. A further comparison between HYMOD, PCEs, and reduced PCEs will also be performed in terms of accuracy and efficiency. NSE will be used for calibration and validation against observed data in order to test the credibility of reduced PCEs for capturing hydrologic behaviors of the Xiangxi River watershed.

This paper is organized as follows. Section 2 reviews basic concepts of PCE and PCM. Section 3 introduces the proposed methodology. Section 4 presents a case study of the conceptual hydrologic model applied to the Xiangxi River watershed, China. In this section, detailed discussions of results and comparisons of various methods are presented. Finally, conclusions are presented in section 5.

2. Polynomial chaos and probabilistic collocation methods

The PCE, introduced by Wiener (1938), is a powerful tool for characterizing uncertainties in model outputs through a series expansion of Hermite polynomials in terms of standard random variables. Consider a model $y_t = f(\theta, x_t, u_t)$ where x_t is the (internal) state vector, u_t is the input vector, and θ represents the vector of time-invariant model parameters. When model parameters are considered as random variables, y_t represents model outputs from hydrologic simulations in a probabilistic space, and can be approximated as multivariate orthogonal polynomials. The vector θ can be expressed as $T[\xi(\theta)]$, where $\xi(\theta) = [\xi_1(\theta_1), \ldots, \xi_n(\theta_n)]^T$ is a vector of standard normal random variables with zero mean and unit variance. To standardize the representation of random variables, a variety of transformations in terms of standard normal random variables are presented by Isukapalli (1999). Then the hydrologic model output can be approximated by a PCE, given by

$$y_t(\theta, x_t, u_t) = a_{0t}(x_t, u_t) + \sum_{i_1=1}^{\infty} a_{i_1 t}(x_t, u_t)\Gamma_1\left(\xi_{i_1}(\theta_{i_1})\right) + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1} a_{i_1 i_2 t}(x_t, u_t)\Gamma_2\left(\xi_{i_1}(\theta_{i_1}), \xi_{i_2}(\theta_{i_2})\right) + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2} a_{i_1 i_2 i_3 t}(x_t, u_t)\Gamma_3(\xi_{i_1}(\theta_{i_1}), \xi_{i_2}(\theta_{i_2}), \xi_{i_3}(\theta_{i_3})) + \ldots, \quad (1)$$