Reliability Engineering and System Safety 169 (2018) 137–148

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Incomplete statistical information limits the utility of high-order polynomial chaos expansions

Sergey Oladyshkin*, Wolfgang Nowak

Department of Stochastic Simulation and Safety Research for Hydrosystems (IWS/SC SimTech), University of Stuttgart, Pfaffenwaldring 5a, Stuttgart 70569, Germany

A R T I C L E I N F O

Keywords:
Polynomial chaos expansion
Orthonormal basis
Uncertainty quantification
Incomplete statistics
Adequate expansion order

A B S T R A C T

Polynomial chaos expansion (PCE) is a well-established massive stochastic model reduction technique that approximates the dependence of model output on uncertain input parameters. In many practical situations, only incomplete and inaccurate statistical knowledge on uncertain input parameters are available. Fortunately, to construct a finite-order expansion, only some partial information on the probability measure is required that can be simply represented by a finite number of statistical moments. Such situations, however, trigger the question to what degree higher-order statistical moments of input data are increasingly uncertain. On the one hand, increasing uncertainty in higher moments will lead to increasing inaccuracy in the corresponding chaos expansion and its result. On the other hand, the degree of expansion should adequately reflect the non-linearity of the analyzed model to minimize the approximation error of the expansion. Observation of the PCE convergence when statistical input information is incomplete demonstrates that higher-order PCE expansions without adequate data support are useless. Moreover, it makes apparent that PCE of a certain order is adequate just for a corresponding amount of available input data. The key idea of the current work is to align the order of expansion with a compromise between the degree of non-linearity of the model and the reliability of statistical information on the input parameters. To assure an optimal choice of the expansion order, we offer a simple relation that helps to align available input statistical data with an adequate expansion order. As fundamental steps into this direction, we propose overall error estimates for the statistical type of error that results from inaccurate statistical information plus the error that results from truncating the expansion of a non-linear model. Our key message is that any order of expansion is only justified if accompanied by reliable statistical information on input moments of a certain higher order.

1. Introduction

Research over several decades has shown that modeling plays a very important role in reconstructing (as far as possible) the complex picture of natural systems and offers a unique way to predict behaviors of the multifaceted processes at play in such complex systems. Most physical processes appearing in nature are non-linear and, as a consequence, the required mathematical models are non-linear as well. Additionally, many natural systems are plagued in modeling by the ubiquitous presence of uncertainty. The influence of uncertainties onto predictions of system behavior is often so strong that it may become the dominant aspect in simulations for applied tasks [1]. Moreover, modern simulation models often demand considerably extensive computational power that makes traditional approaches for stochastic simulation (e.g. Monte Carlo simulation [2] and related approaches [3]) almost impossible. The greatest challenge of the overall modelling procedure is to construct reliable and feasible models that can adequately describe underlying physical concepts and, at the same time, account for uncertainty. A reasonably fast and attractive approach to quantify uncertainty in complex and non-linear systems is to approximate the uncertain model prediction through the polynomial chaos expansion (PCE).

Polynomial chaos expansion. Polynomial chaos expansion is an efficient approach that offers a massive reduction of computational costs in uncertainty quantification. PCE was originally introduced by Wiener [4] in 1938. The key idea of chaos expansion theory consists in projecting a full-complexity model onto orthogonal or orthonormal polynomial bases over the parameter space. Such a reduction of an original model allows to capture the non-linear dependence of quantities of interest on uncertain input parameters [5]. During the last decades, PCE technique has increased in popularity for different applications [6–9]. PCE has been combined with sparse integration rules [10–12] and optimal sampling rule has been proposed [13]. The adaptive multi-element polynomial

* Corresponding author.
E-mail address: Sergey.Oladyshkin@iws.uni-stuttgart.de (S. Oladyshkin).

http://dx.doi.org/10.1016/j.ress.2017.08.010
Received 14 October 2016; Received in revised form 3 August 2017; Accepted 10 August 2017
Available online 12 August 2017

chaos approach [14] has been used to assure flexibility in treating the input distribution. Recently, an attempt to overcome problems with physically constraint variables in PCE was suggested [15]. Thanks to the so-called non-intrusive approach to PCE [16–18], which does not require manipulation of the governing equations, PCE has even been applied to complex physical systems [19–21]. The expanded model which is expressed in the form of polynomials can be evaluated extremely quickly and even in real-time. This particular feature of PCE is highly valuable for applied realistic problems and hence a variety of tasks related to uncertainty quantification could be accomplished at reasonable computational costs. The paper [22] showed how to use PCE for robust design under uncertainty with controlled failure probability. Recently, sensitivity analysis based on PCE decomposition [23–25] and [26] has received increased attention. The papers [27–29] and [30] demonstrate how classical PCE can deliver the information required for global sensitivity analysis at low computational costs. Due to feasibility of PCE, optimization tasks can be performed for computationally demanding applications, see e.g. [31] and [32]. Additionally, a stochastic model calibration framework was developed [33–35] based on strict Bayesian principles combined with the PCE.

Incomplete and inaccurate statistical input data. Statistical input data plays a very important role in the overall modeling process. Unfortunately, statistical information about the uncertain input data is very limited in realistic applications and very often is subjective, incomplete or inaccurate. Assuming the completeness of input statistical information can be acceptable for testing or research purposes, but when the goal is to tackle realistic applied problems, the used methods should be able to deal with incomplete statistical input. Treating incomplete information on input statistics is very challenging because it produces uncertainty in statistical quantities of model outputs [36]. This fact has been recognized in the area of reliability engineering a long time ago [37]. The study [1] illustrates that errors or additional subjective assumptions in data interpretation can severely bias uncertainty quantification and risk assessment. Thus, output statics can be unreliable and uncertain [38]. The paper [39] suggests to build confidence intervals by bootstrap resampling in order to evaluate the reliability of the sensitivity indices constructed via the PCE. As often encounter in practice, limited samples of raw data can be an incomplete source of statistical information because sample data sets do not contain perfect or complete information on the probability distribution of model input parameters. The incomplete knowledge of the statistical input directly translates to uncertainty in specifying its probability measures. The chaos expansion (among other approaches) can be as well applied to situations where only incomplete and inaccurate statistical knowledge on uncertain parameters is available, but it will yield a biased estimate of the output statistics. However, it has been shown [40] that PCE at finite expansion order only demands existence and knowledge of a finite number of statistical moments for the input parameters and does not require the complete knowledge or even existence of a probability measure. Such circumstances trigger the question to what degree higher-order statistical moments of model input are increasingly uncertain. On the one hand, increasing uncertainty in higher moments will lead to increasing inaccuracy in the corresponding chaos expansion and its result. On the other hand, the degree of expansion should adequately reflect the non-linearity of the analyzed model to minimize the approximation error of the expansion. Therefore, the main challenge is to understand better which order of expansion is adequate for a corresponding amount of available statistical input information. The paper [40] provides a data-driven formulation of the PCE and already contains an implicit relation between statistical moments and the expansion terms. This data-driven formulation introduced as arbitrary polynomial chaos expansion has already been exploited for applications in various disciplines (see e.g., [41–44]). However, it is not yet very apparent that blindly increasing the expansion order could be not beneficial for the applied problems where only limited data is available for the analysis. Thus, the data-driven idea introduces in the paper [40] could be extended to understand how incomplete statistical information limits the utility of high-order PCEs.

Approach and contributions. The scope of the current work is to align the order of expansion with a compromise between the degree of non-linearity of the model and the reliability of statistical information on the input parameters. We would like to emphasize that any order of expansion is only justified if it is accompanied by reliable statistical information on input moments of a certain higher order. In Section 2 we deliver the necessary mathematical material to construct PCE from raw statistical moments. Moreover, in Section 3 we offer an inverse solution to the procedure presented in Section 2 and we demonstrate how raw moments can be reconstructed from an existing orthogonal polynomial basis. Thus, we show that the statistical moments characterizing a probability measure are the only source of information that is propagated in all polynomial expansion-based stochastic approaches. Additionally, it makes transparent that only a finite number of moments has to be taken into consideration for constructing a finite-order expansion. Using a simple analytical model, we show in Section 4 how strong uncertainty in input statistical moments could be propagated onto PCE expansion coefficients. Additionally, Section 4 offer an explicit form for the projection of monomials onto a polynomial basis which forms the basis to investigate the sensitivity of PCE to statistical input moments for different degrees of model non-linearity. To assess the impact of incomplete statistical input information in Section 5, we perform a robustness analysis and illustrate the convergence of PCE under incomplete input statistics that are expressed through limited data samples. Furthermore, Section 5 proposes to align the complexity level with the reliability level of statistical information on the input parameters and offers a very simple relation that puts into correspondence available input statistical data with an adequate expansion order. Thus, Sections 2 – 4 offer necessary developments used to align the expansion order with input statistical information and Section 5 delivers the main achievements that are relevant for practical applications.

2. Constructing PCE from raw statistical moments

Section 2 offers the necessary mathematical material to construct PCE from the raw statistical moments via the Hankel matrix of moments. In that way, we will demonstrate that only a limited amount of statistical input information will be employed for construction of a finite order expansion. Please note that we use the Hankel matrix of moments only for analytical reasons and do not advise this approach for actual computations. A recent study [45] has provided numerical evidence that emphasizes the well-known fact that the Hankel matrix often has a poor numerical condition, and other approaches perform more accurately in numerical practice.

2.1. Arbitrary polynomial chaos expansion

We will consider a random process in the probability space (Ω, A, Γ) with space of events Ω, σ-algebra A and probability measure Γ, see e.g. [46] for details. Let us denote a model as $Y(\xi)$ with model input $\xi \in \Omega$ and model output Y. The model $Y(\xi)$ can be represented in differential, integral or closed analytical form. Since ξ is a random variable, so is Y. According to the theory of PCE introduced by [4], the random variable $Y(\xi)$ can be expanded in ξ and approximated in the following manner:

$$Y(\xi) \approx \sum_{i=0}^{d} c_i P^{(i)}(\xi), \tag{1}$$

where d is the order of expansion and $\left\{P^{(0)}, \dots, P^{(d)}\right\}$ forms a polynomial basis that is orthogonal (or orthonormal) with respect to Γ. That order is typically found as a compromise between computational costs and the non-linearity of the underlying model $Y(\xi)$.