EFFICIENCY OF UNCERTAINTY PROPAGATION METHODS FOR ESTIMATING OUTPUT MOMENTS

Samira Mohammadi, Selen Cremaschi*
Auburn University – Auburn, AL 36830

Abstract

Uncertainty propagation methods are used to estimate the distribution of model outputs resulting from a set of uncertain model outputs. There are a number of uncertainty propagation methods available in literature. This paper compares six non-intrusive uncertainty propagation methods, Latin Hypercube Sampling, Full Factorial Integration, Univariate Dimension Reduction, Halton series, Sobol series, and Polynomial Chaos Expansion, in terms of their efficiency for estimating the first four moments of the output distribution using computational experiments. The results suggest employing FFNI if there are few uncertain inputs, up to three. Uncertainty propagation methods that utilize Halton and Sobol series are found to be robust for estimating output moments as the number of uncertain inputs increased. In general, higher order polynomial chaos expansion approximations (3^{rd}-5^{th} order) obtained accurate estimates of model outputs with fewer model evaluations.

Keywords

Uncertainty Propagation Method, Halton and Sobol series, Polynomial Chaos Expansion, Full Factorial Numerical Integration, Univariate Dimension Reduction

Introduction

The steady increase in computational power enabled the widespread utilization of simulation models to assist decision making in chemical and energy process designs. However, there are many sources of uncertainty in these models, e.g., in the model inputs and parameters. This leads to uncertainties in the outputs of the models, which in turn may influence key decisions on the design and operations of the processes and associated integrated equipment. Efficient and effective approaches are required to qualitatively characterize the uncertainty of model outputs, which, as a consequence, will increase the confidence in the predictions of these models, and will enable robust design and operation of the systems.

In general, uncertainty analysis can be carried out in three steps: (i) identify sources and types of uncertainty in the model, (ii) select appropriate mathematical representations for these uncertainties, and (iii) choose and apply an appropriate propagation method to quantify the resulting uncertainty. This paper considers the third component, specifically on uncertainty propagation methods (UPMs).

Uncertainty propagation methods are used where a function $g(\boldsymbol{X})$ is needed to predict the value of a dependent variable y, where $g(\boldsymbol{X})$ may be an analytical function, or a black-box model (Lee and Chen, 2009). Here, y is dependent on $\boldsymbol{X}$, with $\boldsymbol{X}$ representing a $d \times 1$ vector whose components contain the values of the independent variables, and d the number of independent variables. In this case, at least one of the components of $\boldsymbol{X}$ is a random variable with a probability density function or marginal distribution function and correlations (Lee and Chen, 2009). The propagation of uncertainties in $\boldsymbol{X}$ can be carried out by intrusive or non-intrusive UPMs. While intrusive UPMs require the modification of the original model, non-intrusive UPMs, focus of this paper, treat the model as a black-box and hence have much wider applicability Simulation-based UPMs are popular and have been used in many applications because of their simplicity and robustness (Roy and

* To whom all correspondence should be addressed

Oberkampf, 2011). Examples of these methods include the Monte Carlo simulation method (Dieck, 2007), Halton series (Wong et al., 1997), Sobol sequences (Sobol, 1967). There are also UPMs that use surrogate models, such as polynomial chaos expansion (Ghanem and Spanos, 1982). However, little work focuses on systematically comparing the efficiencies of UPMs for black-box models with different characteristics. Existing studies generally compare a limited number of UPMs (e.g., Burhenne and et al., 2011; Garud and et al., 2017). Lee and Chen (2009) and Fahmi and Cremaschi (2016) compared a large set of non-intrusive UPMs in terms of their ability to estimate output distribution moments. However, these studies fixed the number of black-box function evaluations in their analysis, and hence, do not provide any insight on how different UPMs compare to each other as the number of function evaluations change. For especially complex black-box models, such as computational fluid dynamics simulations, a single model evaluation can be computationally expensive. For such models, selection of a UPM that provides accurate estimates of output uncertainty with fewer function evaluations becomes important.

This paper computationally investigates the efficiencies of non-intrusive UPMs in terms of their ability to estimate the moments of the output distributions. We use a number of functions with varying degrees of nonlinearity and with different number of uncertain inputs. The methods considered are Latin Hypercube Sampling (LHS), Full Factorial Integration (FFNI), Univariate Dimension Reduction (UDR), Halton series, Sobol series, and Polynomial Chaos Expansion (PCE). The estimates of the first four statistical moments are used for comparisons.

Uncertainty propagation methods in the analysis

In Latin Hypercube Sampling (LHS), each uncertain-parameter range is partitioned into equal probability bins, where number of bins is equal to the number of required samples, n. A random sample for each uncertain parameter is drawn from each of these bins. Then, a random sample is selected for each uncertain parameter to construct n samples (McKay et al., 1979). Because of the initial partitioning, LHS generally provides good coverage of each uncertain parameter space (McKay et al., 1979). However, because the pairing of the sample points between parameters are done randomly, the coverage of multivariable space may suffer. Furthermore, because the number of required samples determines the number of initial bins the random samples are drawn from, LHS cannot be used in a sequential sampling scheme (Nuchitprasittichai & Cremaschi, 2012).

Full factorial numerical integration (FFNI) and univariate dimension reduction (UDR) are numerical integration methods. Thus, they provide an estimate of output distribution moments via direct numerical integration. First, each uncertain parameter is discretized into m number of nodes. Given m, the numerical integration weights (w_{j_n}) and the corresponding sampling points for each uncertain variable (x_{j_n}) are obtained from Gauss-Hermite, Gauss-Legendre, or Gauss-Laguerre equation depending on the distribution of the uncertain variable (Lee & Chen, 2009). The k^{th} statistical moment of the output distribution, (i.e., $E(y^k)$ where $y = g(\boldsymbol{X})$) is estimated using an appropriate quadrature formula as shown in Eq. 1.

$$E(y^k) = \sum_{j_1=1}^{m} w_{j_1} \dots \sum_{j_n=1}^{m} w_{j_n} \times [g(x_{j_1} \dots x_{j_n})]^k \quad (1)$$

Both FFNI and UDR use Eq. 1 to estimate the output distribution moments. However, FFNI enumerates all possible combinations of x_{j_n} and calculates the corresponding $g(\boldsymbol{X})$. Assuming there are n uncertain parameters, the required number of $g(\boldsymbol{X})$ evaluations is m^n for m nodes. As the number of uncertain parameters increases, the required number of function evaluations (i.e., sample points) increases exponentially even when for a small number of nodes (e.g., two or three).

UDR approximates the overall multivariate function output as an addition of many univariate functions prior to calculating the statistical moments (Lee & Chen, 2009). This approximation function is given in Eq. 2.

$$g(\boldsymbol{X}) \approx \hat{g}(\boldsymbol{X}) = \sum_{i=1}^{n} g_i(\boldsymbol{X}_i) - (n-1) g(\boldsymbol{\mu}_{\boldsymbol{X}}) \quad (2)$$

The $g_i(\boldsymbol{X}_i)$ in Eq. 2 are computed using the value of one variable from the sampling pool of the $m \times n$-dimensional $\boldsymbol{X}$ while keeping the rest of the variables at their mean values ($\mu_{\boldsymbol{X}}$) (Eq. 3). One more function evaluation is required where all uncertain parameters are at their mean values. With UDR, the required number of function evaluations is $mn + 1$. It is worth noting that UDR is equivalent to FFNI for one dimensional functions.

$$g_i(\boldsymbol{X}_i) = g\left(\boldsymbol{X}_i, \boldsymbol{X}_{i'} = \mu_{\boldsymbol{X}_{i'}}\right) \ \forall i, i' \in \{1,2,\dots,n\}, i \neq i' \quad (3)$$

Halton series is a low-discrepancy quasi-random sequence (Wong et al., 1997). Given m as the size of the sample and n number of uncertain parameters, the coordinates of the j^{th} sample point is defined as $\left(\Phi_{p(1)}(j-1), \Phi_{p(2)}(j-1), \dots, \Phi_{p(n)}(j-1)\right)$ where $p(n)$ is an arbitrarily selected prime number that satisfies $p(1) < p(2) < \dots < p(n)$, and $\Phi_p(j)$ is defined in Eq. 4,

$$\Phi_p(j) = \frac{a_0}{p^1} + \frac{a_1}{p^2} + \frac{a_2}{p^3} + \dots + \frac{a_r}{p^{r+1}} \quad (4)$$

where each a_k is an integer in $[0, p-1]$ that satisfies: $j = a_0 + a_1 p + a_2 p^2 + \dots + a_r p^r$.

Sobol series are also low-discrepancy quasi-random sequences. They are designed to sample uniformly from multi-dimensional spaces (Saltelli et al., 2010). The sample points using Sobol series are generated such that the location of each additional sample is related to the location of the existing sample points, which prevents generations of clusters and gaps (Burhenne et al., 2011). In order to produce the j^{th} component of i^{th} point of samples a primitive polynomial is used (Eq. 5).

$$x^{s_j} + a_{1,j}x^{s_j-1} + a_{2,j}x^{s_j-2} + \dots + a_{s_j-1,j}x + 1 \qquad (5)$$

where s_j is the degree of the polynomial and the coefficients $a_1, \dots, a_{s_j-1,j}$ are either 0 or 1 (Joe and Kuo, 2008). The j^{th} component of i^{th} point is given in Eq. 6.

$$x_{i,j} = i_1 v_{1,j} \oplus i_2 v_{2,j} \oplus \dots \qquad (6)$$

where $\oplus$ is the bit-by-bit exclusive-or operator and $v_{k,j}$ are the direction numbers (Eq. 7),

$$v_{k,j} = \frac{m_{k,j}}{2^k} \qquad (7)$$

and $m_{k,j}$ are sequence of positive integers that satisfy Eq. 8.

$$m_{k,j} = 2a_{1,j}m_{k-1,j} \oplus 2^2 a_{2,j}m_{k-2,j} \oplus \dots \oplus m_{k-s_j,j} \qquad (8)$$

Polynomial chaos expansion (PCE) approximates "well-behaved" random variables as a polynomial series of standard normal random variables (Crestaux et al., 2009; Lee & Chen, 2009). The output of PCE is a random variable expressed as a polynomial series of standard normal random variables, therefore the statistical moments of the resulting output distribution are calculated using these polynomial series (Lee & Chen, 2009). The general form of PCE of a random variable $u(\theta) \in L^2$ can be written as:

$$u(\theta) = a_0\Gamma_0 + \sum_{i=1}^{\infty} a_{i_1}\Gamma_1\left(\xi_{i_1}(\theta)\right) + \sum_{i=1}^{\infty}\sum_{i_2=1}^{i_1} a_{i_1 i_2}\Gamma_2\left(\xi_{i_1}(\theta), \xi_{i_2}(\theta)\right) + \cdots \qquad (9)$$

In Eq. 9, Γ_p are the Hermite polynomials of order p as functions of $\xi_i(\theta)\ \forall i \in \{1,2,\dots\infty\}$ which are the standard normal variables that are viewed as a function of the random event θ, and a_i's are deterministic coefficients (Lee & Chen, 2009). For practical purposes, PCE should be truncated to a polynomial degree p. Eq. 9 can be rewritten in a simpler form and for a polynomial order of p with n uncertain parameters (inputs) as shown in Eq. 10.

$$u(\xi_1, \xi_2, \dots, \xi_n) \approx \sum_{i=0}^{\frac{(p+n)!}{p!n!}-1} b_i\psi_i(\xi(\theta)) \qquad (10)$$

where b_i and $\psi_i(.)$ correspond to $a_{i_1 \dots i_p}$ and Γ_p, respectively. The deterministic coefficients, b_i, can be calculated based on the orthogonality of Hermite polynomials as shown in Eq. 11 (Lee & Chen, 2009). The expected value in the denominator of Eq. 11 can be evaluated analytically, but the value in the numerator has to be evaluated numerically.

$$b_i = \frac{E[u\psi_i(\xi(\theta))]}{E[\psi_i^2(\xi(\theta))]} \qquad (11)$$

If the uncertain parameters are not standard normal random variables, they should be transformed into standard normal random variables prior to using PCE as shown in Eq. 12 (Lee & Chen, 2009).

$$\xi_i = \Phi^{-1}(F(X_i)) \qquad (12)$$

where X_i is the original random variable, $F(.)$ denotes the cumulative distribution function (CDF) of the original random variable, and Φ denotes the CDF of the standard-normal distribution.

Computational Experiments

In computational experiments, UPMs were used to propagate the uncertainty of the input(s) to the output(s) of a set of functions, whose analytical forms are known. The functions include power function with exponents ranging from one to five ($y(x) = x^r\ \forall r \in \{1,2,\dots,5\}$), the Ackley function from one to seven dimensions (d in Eq.13) (Back, 1996), and G function with input dimensions (d in Eq. 14) from one to five (Marrel et al., 2008). All inputs were assumed to be uncertain with three different distributions, Uniform (0, 10), Normal (5, 3), Lognormal (1.5, 0.37) except for G function, where all input parameters were uniformly distributed (Uniform (0, 1)) (Marrel et al., 2008).

$$Ackley_d(x) = -20\exp\left(-0.2\sqrt{\frac{1}{d}\times\sum_{i=1}^{d}x_i^2}\right) - \exp\left(\frac{1}{d}\sum_{i=1}^{d}\cos(2\pi x_i)\right) + 20 + \exp(1) \qquad (13)$$

$$G_d(x) = \prod_{i=1}^{d}\frac{|4x_i - 2| + a_i}{1 + a_i}, \ where\ a_i = \frac{i-2}{2} \qquad (14)$$

The computational experiments started with the minimum number of function evaluations required to calculate estimates of the first four moments of the output for each UPM. Then, the number of function evaluations were increased, and moment estimates were calculated until the number of function evaluations reached 1×10^6. The experiments included PCE polynomials truncated to four different orders, i.e., $p = \{2,3,4,5\}$. We used function evaluations generated according to Sobol and Halton series, and FFNI for the numerical integration (denominator of Eq. 11). The performance of the UPMs were assessed based on the quality of the first four statistical moments of the output distribution using the minimum number of function evaluations required for a statistical moment to reach and remain inside a band whose width is equal to a pre-determined percentage of the true moment value. The 'true' values of the moments were obtained with Monte Carlo simulation with 5×10^6 function evaluations. Experiments were coded in Python 3.6, and packages Sobol_seq (Naught101, 2017) and Chaospy (Feinberg and Langtangen, 2015) were utilized for implementing UPMs using Sobol series and polynomial chaos expansion, respectively.Results and Discussion

Results for the power function

The results for the power function are summarized in Fig. 1, where x-axis is the exponent and y-axis is the minimum number of function evaluations for the statistical moments to reach and remain within a 2% band of their true values. In figures, the order of the PCE and method used for numerical integration in PCE is specified by the following convention: "PCEorder-integration approach", where S, H, and F correspond to Sobol and Halton series, and FFNI.

Comparison of the plots on the first row of Fig. 1 reveals that the minimum numbers of function evaluations for Sobol and Halton series and LHS for mean estimates become considerably high as the exponent increases. A similar trend is observed for PCEs with orders $p = \{2,3\}$. The mean estimates of FFNI and PCE with orders $p = \{4,5\}$ settle to the true values quickly with few function evaluations even for higher exponents. Similar conclusions can be reached for the minimum number of function evaluations required for estimation of standard deviation (second row in Fig. 1), skewness (third row in Fig. 1), and kurtosis (fourth row in Fig. 1). In general, FFNI and higher order PCEs require fewer function evaluations than Sobol and Halton series, LHS, and low order PCE, for which the

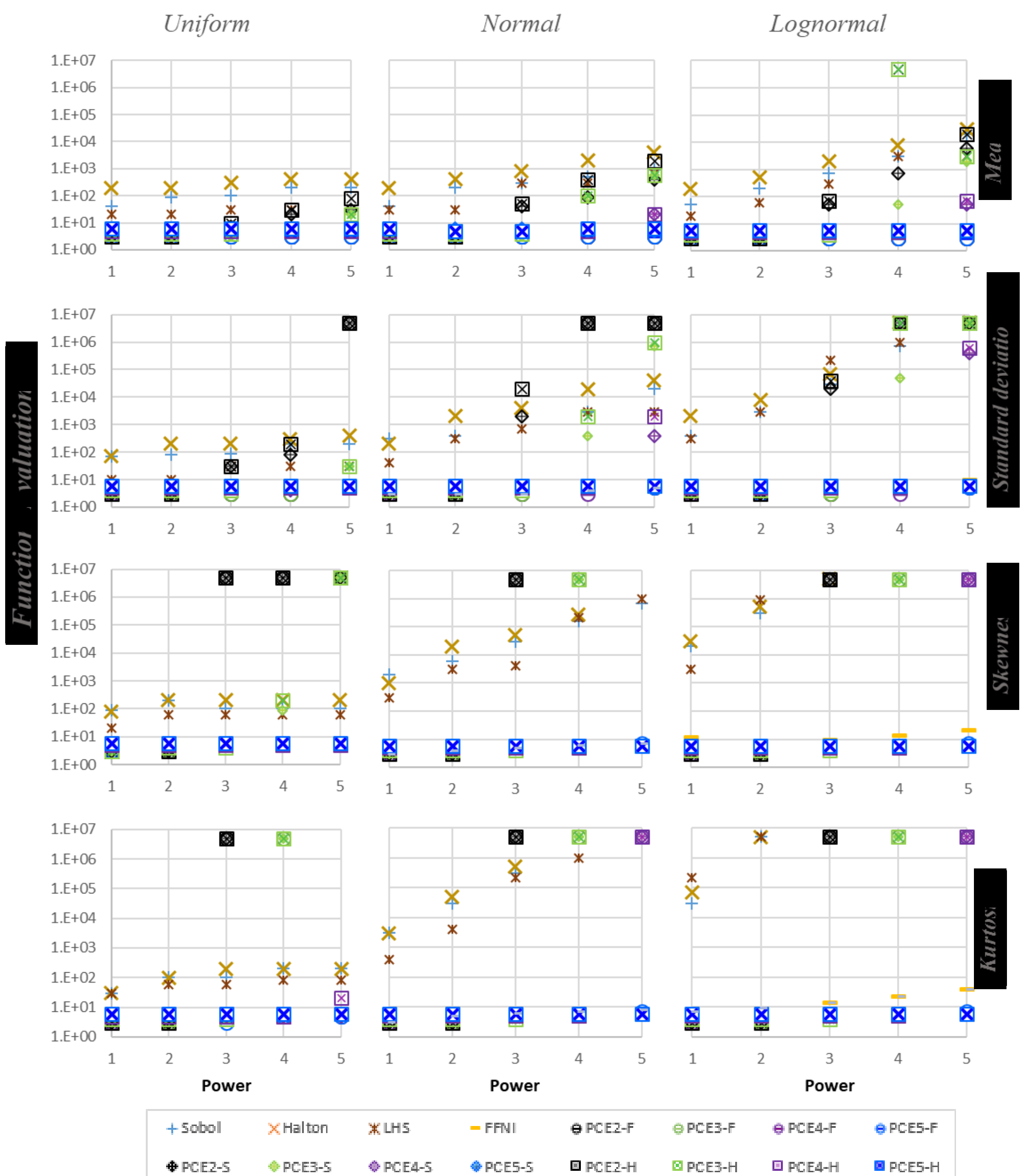

Figure 1. The minimum number of function evaluations for estimating mean, standard deviation, skewness, and kurtosis within 2% of their true values for the power function, where the input is distributed uniformly, normally, and lognormally

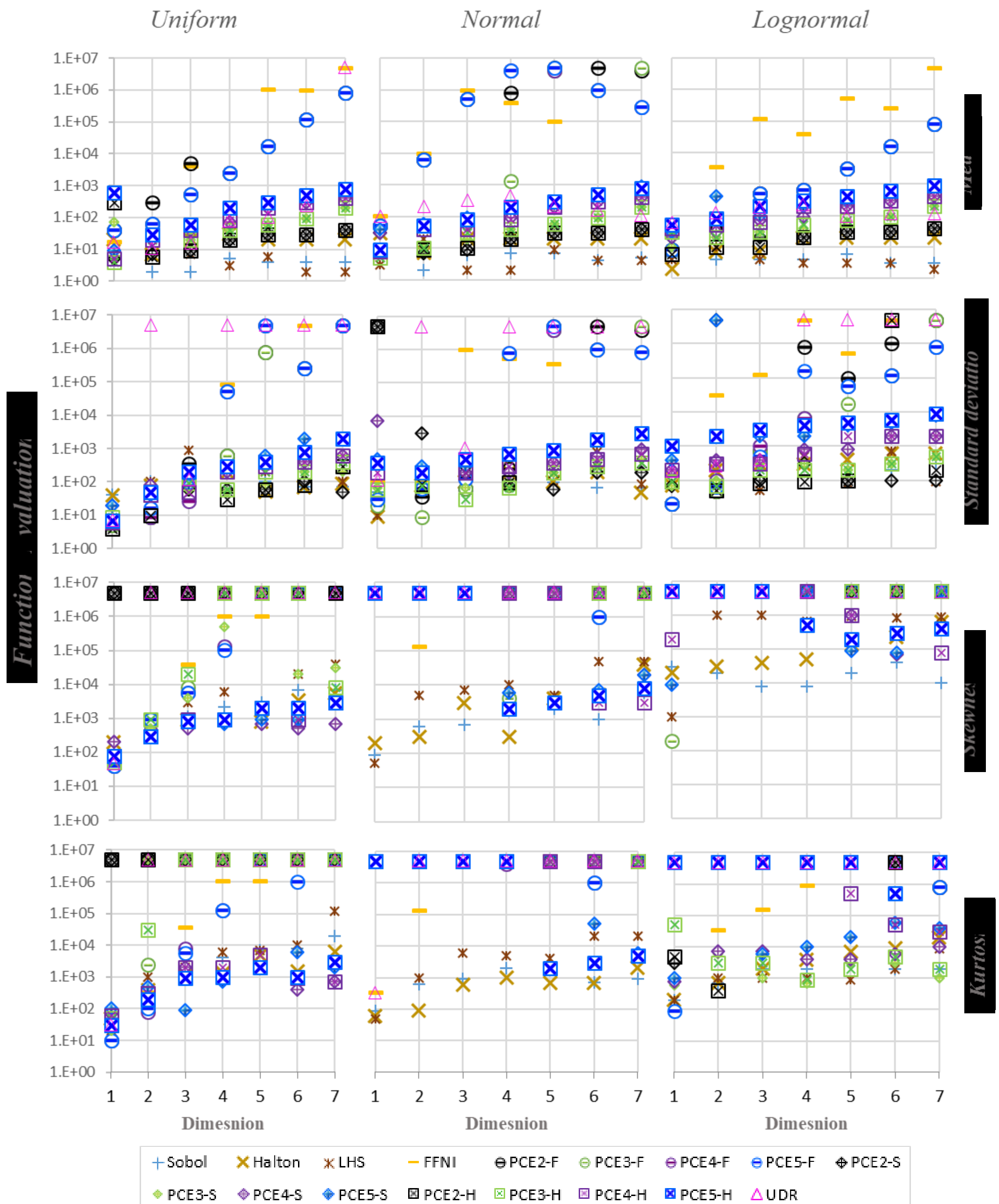

Figure 2. The minimum number of function evaluations for estimating mean, standard deviation, skewness, and kurtosis within 5% of their true values for Ackley function, where the input is distributed uniformly, normally, and lognormally

minimum number of function evaluations grow almost exponentially as the exponent increases.

Results for the Ackley and G functions

Figures 2 and 3 summarize the results for the Ackley and G functions, respectively. The width of the band is 5% for the plots. Figures 2 and 3 reveal that the minimum number of function evaluations increases gradually for all moment estimations as the number of uncertain inputs increases for all UPMs. However, in general, Sobol and Halton series and LHS are more robust compared to the other methods (Figs. 2 and 3). As the number of uncertain inputs increases, the minimum number of function evaluations increases quickly for FFNI and PCE with FFNI as the numerical integrator. The increase in the minimum number of function evaluations with the number of uncertain inputs becomes steeper for higher order moment estimations.

It is worth noting that UDR generated poor estimates for the G function due to the characteristics of the function, product of the input variables with mean value of 0.5, and hence, the UDR results is not included in Fig. 3.

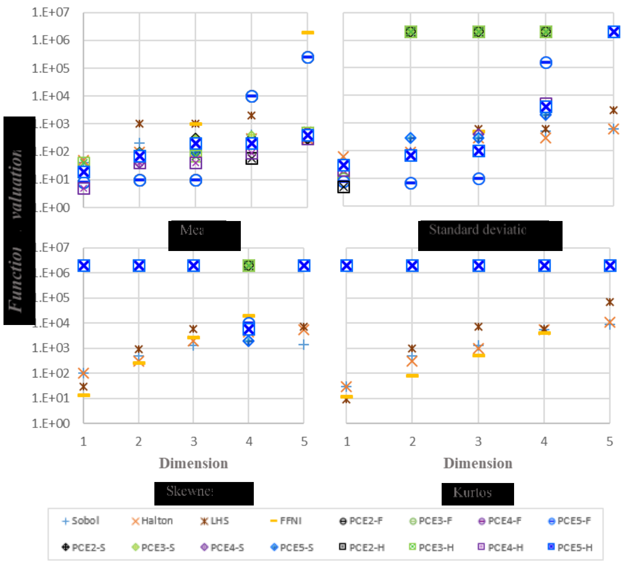

Figure 3. The minimum number of function evaluations for estimating mean, standard deviation, skewness, and kurtosis within 5% of their true values for G function

Impact of input distributions

The impact of input distributions can be deducted by column-wise comparison of the plots in Figs. 1 and 2. In average, the minimum number of function evaluations increases as the input distribution changes from uniform to normal and from normal to lognormal. This trend is more pronounced for higher exponent values and for Sobol and Halton series, LHS, lower order PCEs, and PCEs that utilize Sobol and Halton series for numerical integration in Fig. 1.

Conclusions

This paper compared the performance of six nonintrusive uncertainty propagation methods in estimating the first four moments of output distribution using computational experiments. The methods considered are Latin Hypercube Sampling (LHS), Full Factorial Integration (FFNI), Univariate Dimension Reduction (UDR), Halton series, Sobol series, and Polynomial Chaos Expansion (PCE). The results suggest that Sobol and Halton series, and LHS may not be appropriate uncertainty propagation methods for models with uncertain inputs in highly nonlinear relationships. However, they are quite robust for high numbers of uncertain inputs. PCEs with low order polynomials may not be the proper choice for highly non-linear models, however they are stable as the number of uncertain inputs increases, and vice versa is true for high order PCEs. The FFNI and PCE that uses FFNI require computationally prohibitive number of function evaluations for more than three or four uncertain inputs.

Acknowledgments

This work was funded by NSF grant #1743445 and RAPID Manufacturing Institute, the U.S.A. The work was completed in part with resources provided by the Auburn University Hopper Cluster.

References

Back, T. (1996). Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms. New York: Oxford University Press.

Burhenne.S., Jacob.D and Henze G.P. (2011), Sampling Based on Sobol‟ Sequences for Monte Carlo Techniques Applied to Building Simulations, 12th Conference of International Building Performance Simulation Association, Sydney

Crestaux, T., Le Maˆıtre, O., & Martinez, J.-M. (2009). Polynomial chaos expansion for sensitivity analysis. Reliability Engineering & System Safety, 94, 1161-1172.

Dieck RH. Measurement uncertainty: methods and applications. 4th edition.

Fahmi, I. & Cremaschi, S. (2016). Computational Experiments on Sampling Methods for Uncertainty Propagation and the Implications for Simulation-Based Optimization. 10.1016/B978-0-444-63428-3.50301-5.

Feinberg F., Langtangen H.P. (2015). Chaospy: An open source tool for designing methods of uncertainty quantification, Journal of Computational Science, 11, 46-57.

Garud, S.S., Karimi, I.A., Kraft, M., 2017. Design of computer experiments: a review.

Ghanem RG, Spanos PD (1991) Stochastic finite elements: a spectral approach. Springer, New York

Herzog M., Gilg A., Paffrath M., Rentrop P., Wever U. (2008) Intrusive versus Non-Intrusive Methods for Stochastic Finite Elements. In: Breitner M.H., Denk G., Rentrop P. (eds) From Nano to Space. Springer, Berlin, Heidelberg

Joe, S., Kuo, F.Y. 2008. Notes on generating Sobol sequences, ACM Transactions on Mathematical Software (TOMS), 29(1), 49-57.

Lee, S., & Chen, W. (2009). A comparative study of uncertainty propagation methods for black-box-type problems. Structural and Multidisciplinary Optimization, 37, 239-253.

Marrel, A., Iooss, B., Van Dorpe, F., & Volkova, E. (2008). An efficient methodology for modeling complex computer codes with Gaussian processes. Computational Statistics & Data Analysis, 52(10), 4731-4744.

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics, 21, 239-245.

Naught101. (2017). Sobol sequence implementation. GitHub repository. https://github.com/naught101/sobol_seq.

Nuchitprasittichai, A., & Cremaschi, S. (2012). An algorithm to determine sample sizes for optimization with artificial neural networks. AIChE Journal.

Roy, Christopher & L. Oberkampf, William. (2011). A Complete Framework for Verification, Validation, and Uncertainty Quantification in Scientific Computing (Invited). Computer Methods in Applied Mechanics and Engineering - COMPUT METHOD APPL MECH ENG. 200. 2131-2144. 10.1016/j.cma.2011.03.016.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Computer Physics Communications, 181(2):259–270.

Sobol,I.M. (1967), "Distribution of Points in a Cube and Approximate Evaluation of Integrals". Zh. Vych. Mat. Mat. Fiz. 7: 784–802 (in Russian); U.S.S.R Comput. Maths. Math. Phys. 7: 86–112 Research Triangle Park (NC): ISA; 2007.