Efficiency of uncertainty propagation methods for moment estimation of uncertain model outputs

Samira Mohammadi, Selen Cremaschi*

Department of Chemical Engineering, Auburn University, Auburn, AL 36849, United States

*Corresponding Author: selen-cremaschi@auburn.edu

Abstract

Uncertainty quantification and propagation play a crucial role in designing and operating chemical processes. This study computationally evaluates the performance of commonly used uncertainty propagation methods based on their ability to estimate the first four statistical moments of model outputs with uncertain inputs. The metric used to assess the performance is the minimum number of model evaluations required to reach a certain confidence level for the moment estimates. The methods considered include Monte-Carlo simulation, numerical integration, and expansion-based methods. The '*true*' values of the moments were calculated by high-density sampling with Monte-Carlo simulations. Ninety-five functions with different characteristics were used in the computational experiments. The results reveal that, despite their accuracy, numerical integration methods' performance deteriorates quickly with increases in the number of uncertain inputs. The Monte-Carlo simulation methods converge to the moments' *true* values with the minimum number of model evaluations if model characteristics are not considered or known.

Keywords: Extrinsic Uncertainty, Uncertainty propagation, Moment estimation, Monte Carlo simulation, Numerical integration, Functional expansion

Nomenclature

Symbol	Description
Y	Model output
X_i	i^{th} uncertain input
$g(X)$	Model
μ_i	i^{th} statistical moment
m	Number of samples
n	Number of uncertain inputs
δ	Number of nodes in each dimension
s_j	Order of primitive polynomial for j^{th} component of the Sobol series
$a_{1,j} \dots a_{s_j-1,j}$	Binary coefficients of primitive polynomial for j^{th} component of the Sobol series
β_k	k^{th}digit from right when β is written in binary
$v_{k,j}$	Direction number for k^{th} component of the Sobol series
$p(\mathrm{k})$	k^{th} selected arbitrary prime number
$t_0 \dots t_r$	Coefficients of expansions via prime number $p(\mathrm{k})$ with a maximum order of r
$f(X)$	The joint density function of uncertain inputs
w_{j_k}	Weight of k^{th} point for j^{th} component of the quadrature in full factorial numerical integration
$g_j(X_j)$	j^{th} univariate function
$\hat{g}(X)$	Approximated model

$M_{X_{j'}}$	Mean value of uncertain input $X_{j'}$
$g(M_X)$	Model with uncertain inputs fixed at their average value
p_s	Number of possible sample points for sparse grid method
$U_1^{i_\alpha}$	${i_\alpha}^{th}$ one dimensional quadrature
w_l	Weight of l^{th} sample point for sparse grid method
Γ_p	Orthogonal polynomial of order p
c_i	i^{th} coefficient of orthogonal polynomial
$\xi_i(\theta)$	Standard variable
ψ_i	i^{th} component of the orthogonal polynomial

1. Introduction

With advances in computing systems and improved computational power, simulation models have become popular methods for assisting with decision-making in chemical process design and operation. Many uncertainties present in the simulation models, e.g., in the model inputs and/or parameters, model formulations, and numerical calculations, cause their outputs to be uncertain. In recent years, the uncertainty due to numerical calculations has reduced significantly with advanced computational power (Hüllen et al., 2019). Therefore, the primary sources of uncertainty in simulation outputs are uncertain inputs, uncertain parameters, and model form uncertainty. This study investigates the uncertainty of simulation model outputs due to extrinsic uncertainty, which is the uncertainty resulting from a predefined number of uncertain inputs with known distribution parameters (Ankenman et al., 2008).

Model uncertainty is studied and characterized using the uncertainty quantification (UQ) methods. The UQ methods are also used to reduce uncertainties in the systems to generate reliable output values and increase confidence in the models (Miller et al., 2014). Important steps of UQ are 1) identification of uncertainty sources, 2) characterization of the sources, 3) Uncertainty Propagation (UP), and 4) analyzing

the uncertainties (Gel et al., 2013). Uncertainty propagation investigates the contribution of uncertain sources to the final uncertainty of the model. When only extrinsic uncertainty is considered, the UP methods propagate the uncertainty of the inputs (X) to the model outputs ($Y=g(X)$) of the model $g(.)$ (Lee and Chen, 2009). For propagating extrinsic uncertainty to outputs, UP methods first require selecting the appropriate statistical representation for the uncertain input variables. Next, the UP is carried out for the model to make statistical inferences regarding the outputs. Statistical inferences regarding the uncertain outputs are generally carried out through estimating three main statistical concepts: the probability density function of the outputs, statistical moments of the outputs, and the probability of a certain outcome, such as failure, based on output distribution (Yang et al., 2017).

There are many challenges in UQ and UP, such as discontinuous response surfaces, selection of significant uncertain parameters for models with high dimensionality, highly complex physical/simulation models, and computational cost associated with UP. There are many UP methods in the literature addressing parts of these challenges. (Groen et al., 2014; Luo and Yang, 2017; Wang and Sheen, 2015).

Uncertainty propagation methods are divided into two groups, intrusive and non-intrusive methods. In intrusive methods, the model formulation is needed and modified to propagate input uncertainty. The models are treated as black boxes for non-intrusive methods. Lee and Chen (2009) categorized the non-intrusive UP methods into five groups, 1) simulation-based methods, e.g., Monte Carlo (MC) simulations, 2) local expansion based methods, 3) most probable point-based methods, 4) functional expansion–based methods, e.g., polynomial chaos expansion (PCE), and 5) numerical integration-based methods. It has been established that the moment estimates obtained using local expansion-based UP methods are significantly different from the true values for models with high nonlinearities (Jia et al., 2019; Lee and Chen, 2009). Most probable point-based UP methods are typically used for reliability applications and do not provide accurate estimates of higher statistical moments (Arakere et al., 2010; Padulo et al., 2007). In addition to these five categories, response-surface-based methods have been used in recent years, where the models of interest are represented through surrogate models (Murcia et al., 2018; Sofi et al., 2020; Tripathy et al.,

2016). The response-surface-based methods encompass the fourth category, which is functional expansion-based methods.

Most UP methods require the evaluation of complex simulation models and many model runs (Liu and Gupta, 2007). Carrying out UP for complex or high fidelity models that are computationally expensive to evaluate could be prohibitive for achieving accurate results with some UP methods (Rajabi, 2019). Hence, selecting the appropriate UP method is crucial for efficient and accurate UP.

Multiple studies compared the performance of different UP techniques in terms of accuracy and efficiency to guide selecting an appropriate UP method. Several of these studies compare simulation-based methods to other categories of UP methods. For instance, Klavetter et al. (2012) compared perturbation, Taylor series expansion, and Monte Carlo methods in propagating the uncertainty in slug length and liquid entrainment in gas core to the outputs of a multiphase flow model. The results stated that Taylor series expansion overestimated variance for most outputs, and the other two methods yielded comparable estimates. However, the perturbation method may not provide reasonable uncertainty estimates for models that are not monotonically increasing or decreasing, and it does not provide confidence levels (Klavetter et al., 2012).

Several studies investigated simulation-based methods versus functional expansion-based methods. Safta et al. (2017) and Hunt et al. (2015) compared the accuracy and efficiency of MC simulations, PCE, and Quasi-Monte Carlo (QMC) simulation methods. Both studies concluded that the PCE required fewer model evaluations to converge to the true value of the output mean for the test functions. Aleti et al. (2018) studied the efficiency of MC simulation and PCE methods based on the number of sample points used to estimate the output distribution accurately. The results revealed that the PCE was 90% more efficient than MC methods in terms of the number of numerical calculations.

Jia et al. (2019) evaluated the performance of MC simulation and the numerical integration approaches, including Sparse grid numerical integration (SG), Univariate dimension reduction (UDR), and extended sparse grid methods. They concluded that the SG methods were the most efficient in estimating the first four moments of the output requiring the fewest model evaluations. Allen and Camberos (2009)

compared simulation-based methods to response surface approaches to estimate the output probability density function and calculate the probability of failure, defined as the probability of an event that the output value exceeds a specific critical level, using the probability density function. They employed two models as case studies, one with high nonlinearity and one with high dimension. The results were evaluated based on the number of required model evaluations to predict the desired uncertainty metrics. They concluded that response-surface methods, specially polynomial chaos expansion, accurately estimated the probability of failure with the lowest number of samples compared to other methods. One other conclusion was that accurate output distribution estimates required many samples from the uncertain input space and many model evaluations.

Some studies only considered different simulation-based methods and compared their performances. Both Burhenne et al. (2011) and Hou et al. (2019) studied MC and QMC methods by employing different sampling techniques. The performance was assessed based on accuracy and efficiency in estimating the output mean for a set of test functions in both studies and standard deviation in Hou et al. (2019). The results suggested that QMC methods are efficient and outperform MC methods in most cases.

Other studies investigated the difference in the performance of other UP method categories. Padulo et al. (2007) employed local expansion and most probable point-based approaches in their study. First- and third-order Taylor series expansion and Sigma point methods were used to estimate output uncertainty for four different test functions. Sigma point methods provided better estimates of the output mean and standard deviation for input distributions with high variance. Sigma point methods do not require derivatives, which gives them a computational advantage over Taylor series expansion for functions with expensive derivative calculations. Rajabi (2019) and Tardioli et al. (2016) investigated different response surface-based methods. Rajabi (2019) compared PCE to Gaussian Process Emulation (GPE). The study suggested that although GPE had lower normalized Root Mean Square Error (nRMSE) in estimating the response surface, PCE estimated output mean, standard deviation, and probability density function tails with higher accuracy. In addition, PCE tended to have lower statistical dispersion with noisier input probability distributions. Tardioli et al. (2016) compared PCE, Tchebycheff expansions with sparse grids, kriging (Gaussian process

modeling), and high dimensional model representation (HDMR) methods. The performance was evaluated based on the methods' ability to represent the response surface of the test models at different sample sizes using RMSE as the metric. Tchebycheff expansion was concluded to be efficient due to its use of sparse grids and required a lower number of sample points to get to the desired accuracy. The performance of PCE was observed to be inconsistent, and HDMR provided very close results to the Tchebycheff expansion method requiring a lower number of samples to converge to the desired accuracy for all test models. Compared to the other methods, kriging required a high number of model evaluations and had a higher RMSE for all the case studies.

Two papers compared more than two main categories of UP methods. Lee and Chen (2009) and Fahmi and Cremaschi (2016) included MC, Full Factorial Numerical Integration (FFNI), UDR, and PCE methods in a comparative analysis. Fahmi and Cremaschi (2016) also studied different sampling schemas of random, Halton sequences, and Latin Hypercube sampling (LHS) for both MC and PCE. The number of function evaluations used for the analysis was fixed in both studies. The methods were compared in terms of their ability to estimate the four statistical moments of the model outputs. Lee and Chen (2009) concluded that the performance of the methods depended on the model characteristics, such as nonlinearity and uncertain variable interactions. The results from Fahmi and Cremaschi (2016) revealed that simulation-based methods were more sensitive to existing nonlinearities in the test functions than other methods.

The UP method comparisons carried out in the literature demonstrate the importance of the UP method selection for efficiently propagating the extrinsic uncertainty for obtaining accurate estimates of the output uncertainty and allude to the correlation between the accuracy and efficiency of the UP methods and the model characteristics the method is applied to. The goal of this paper is to establish systematic guidelines for selecting an accurate and efficient UP method considering several important factors: 1) model characteristics, 2) the number of uncertain input variables, 3) uncertainty distribution of the input variables, and 4) performance evaluation for higher-order moments like skewness and kurtosis. The current comparison literature does not consider one or more of these factors in their analysis, and none provides rules of thumb for selecting an appropriate UP method that will efficiently, i.e., with the lowest number of

model evaluations, estimate the output uncertainty. This study aims to fill this gap by thoroughly evaluating the popular methods considering all key factors. Characterizing the impact of the extrinsic uncertainties on the outputs will result in higher confidence in model predictions, which will aid the efficient and robust design and operation of the systems these models represent.

In this study, we compare seven non-intrusive UP methods based on their ability to estimate the first four statistical moments of the outputs of models with uncertain inputs. The methods considered are Monte Carlo simulation using Sobol sequences (Sobol', 1967), Halton series (Halton, 1960), and LHS (McKay et al., 1979), FFNI (Duffy et al., 1998), UDR (Rahman and Xu, 2004), SG (Smolyak, 1963), and PCE (Ghanem and Spanos, 1991). An extensive set of test functions were employed to study the effects of 1) nonlinearities, 2) the number of uncertain inputs, and 3) different input uncertainty distributions for establishing guidelines for selecting efficient UP methods. The efficiency of the methods was evaluated using the minimum number of model evaluations required by each method to converge to a preset gap around the true value of the first four statistical moments. Finally, the guidelines were utilized to determine the most appropriate UP method for two case studies. Section 2 briefly explains the UP methods used in this study. Computational experiments are described in Section 3, followed by results and discussion in Section 4. Section 5 summarizes the concluding remarks and outlines future directions.

2. Uncertainty Propagation Methods

In this study, uncertainty propagation is conducted by estimating the first four statistical moments of the model output $Y = g(X)$ (Figure 1). The distribution for the uncertain input vector X is assumed to be known for each model ($g(.)$). The estimated mean, standard deviation, skewness, and kurtosis are the extracted information of the output uncertainty.

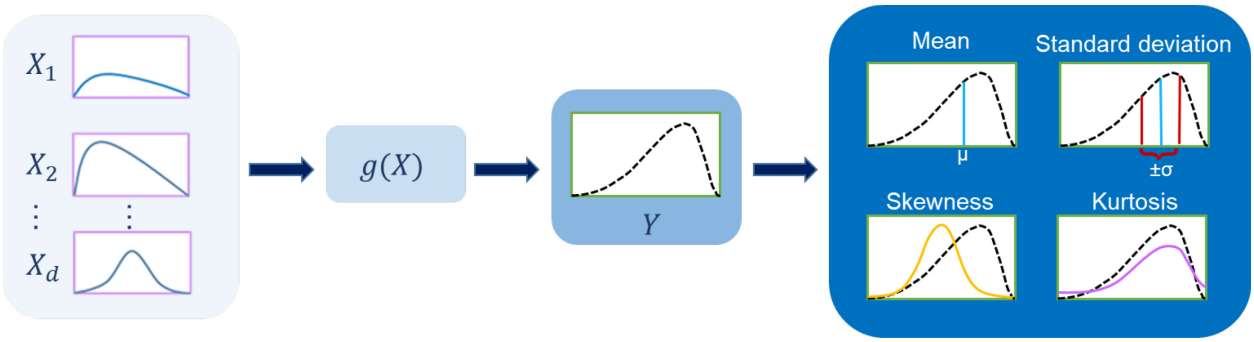


Figure 1. Propagation of uncertain inputs to the simulation output via four statistical moments.

2.1. Monte Carlo simulation-based methods

For Monte Carlo simulation (MCS) based methods, the method of moments (Hansen, 1982) is used to estimate the statistical moments. The i^{th} moment (μ_i) is calculated by Eq. 1,

$$\mu_i = E\left[g(X)^i\right] \approx \frac{1}{m}\sum_{j=1}^{m} g(X_j)^i \tag{1}$$

where $g(X_j)$ is the model value at j^{th} sample point X_j from input distribution(s), and m is the number of sample points. Three methods, LHS (McKay et al., 1979), Sobol sequences (Sobol', 1967), and Halton series (Halton, 1960), are implemented as sampling techniques for determining uncertain input space sample locations and calculating the corresponding model outputs for propagating input uncertainty. These three sampling methods are space-filling sampling techniques (Crombecq et al., 2011), which spread out the sample points evenly throughout the input space. LHS, Sobol sequence, and Halton series sampling schemes are used to generate samples between zero and one. Then, the samples between zero and one are transferred to the original input distribution space using the appropriate reverse cumulative distribution functions. Moreover, Sobol and Halton series are sequential sampling methods (Hou et al., 2019), allowing previous sample points and model evaluations to be reutilized if additional sample points are collected.

2.1.1. Latin Hypercube Sampling (LHS)

In the LHS method (McKay et al., 1979), the range for each uncertain variable is divided into m bins with equal probability, where m is the number of required sample points. Then, a sample is randomly

selected from each bin for each uncertain variable. Next, the samples of different uncertain variables are randomly matched and result in an $m \times n$ sample matrix, where n is the number of uncertain inputs. The initial binning enables LHS to cover the space of each uncertain parameter better than MCS with random sampling. However, the random matching of the samples from different dimensions could cause clusters in the design space and lead to a poor space-filling attribute. Latin Hypercube sampling is not a sequential sampling method because the binning of the uncertain parameters is a function of the number of sample points, and it changes as the value of m is modified (Fahmi and Cremaschi, 2016).

2.1.2. Sobol sequences

Sobol sequences are low-discrepancy pseudo-random series (Sobol', 1967). These sequences are constructed for generating samples as uniformly as possible over the sampling space (Saltelli et al., 2010). Every new sample point is generated based on the location of the existing points, which helps with avoiding clusters and gaps (Burhenne et al., 2011). According to Joe and Kuo (Joe and Kuo, 2008), for the generation of the j^{th} component of Sobol samples, a primitive polynomial of order s_j, must be selected (Eq. 2).

$$x^{s_j} + a_{1,j}x^{s_j-1} + a_{2,j}x^{s_j-2} + \cdots + a_{s_j-1,j}x + 1 \tag{2}$$

The coefficients $a_{1,j}, a_{2,j}, \ldots, a_{s_j-1,j}$ in Eq. 2 are binary values. The j^{th}component of the β^{th} point in Sobol series, $x_{\beta,j}$, is calculated by Eq. 3,

$$x_{i,j} = \beta_1 v_{1,j} \oplus \beta_2 v_{2,j} \oplus \cdots \tag{3}$$

where β_k is the k^{th} digit from the right when β is written in binary $\beta = (\ldots \beta_3\beta_2\beta_1)_2$ and $\oplus$ is the bit-by-bit exclusive-or operator. The parameters $v_{k,j}$ are called the direction numbers (Eq. 4), where $q_{k,j}$ are positive integers calculated in Eq. 5.

$$v_{k,j} = \frac{q_{k,j}}{2k} \tag{4}$$

$$q_{k,j} = 2a_{1,j}q_{k-1,j} \oplus 2^2a_{2,j}q_{k-2,j} \oplus \ldots \oplus q_{k-s_j,j} \tag{5}$$

2.1.3. *Halton series*

Halton sequences (Halton, 1960) are low-discrepancy series used for sampling. Given n number of uncertain inputs, the i^{th} sample point obtained using Halton series is given by Eq. 6,

$$\left(\Phi_{p(1)}(i-1), \Phi_{p(2)}(i-1), \dots, \Phi_{p(n)}(i-1)\right) \tag{6}$$

where $p(\mathrm{n})$ is a selected arbitrary prime number, which is subject to $p(1) < p(2) < \cdots < p(n-1)$. The variable $\Phi_p(i)$ is defined in Eq. 7 (Wong et al., 2005).

$$\Phi_p(i) = \frac{t_0}{p^1} + \frac{t_1}{p^2} + \frac{t_2}{p^3} + \cdots + \frac{t_r}{p^{r+1}} \tag{7}$$

In Eq. 7, t_r is an integer in $[0, p-1]$ that follows Eq. 8, which shows the expansion of integer i with the maximum order of r, where r is any positive integer value, via prime base p (Wong et al., 2005).

$$i = a_0 + a_1 p + a_2 p^2 + \cdots + a_r p^r \tag{8}$$

2.2. *Numerical integration methods*

The output moments of a model with uncertain inputs are defined in Eq. 9, which gives the integral for the i^{th} moment, μ_i (Grimmett and Stirzaker, 2001). This integral can be computed using an appropriate numerical integration method.

$$\mu_i = \int_{-\infty}^{\infty} g(X)^i f(X) dX \tag{9}$$

In Eq. 9, $f(X)$ is the joint density function of the uncertain input variables, X, for the model $g(X)$. Here, we utilize FFNI, UDR, and SG, which are popular numerical integration methods.

2.2.1. *Full factorial numerical integration*

Full factorial numerical integration (FFNI) employs a weighted sum of the model output values at specified input values (Duffy et al., 1998). For estimating the statistical moments of model output with uncertain inputs, the model is evaluated at $m = \delta^n$ specific input values, which can also be referred to as sample locations, where n is the number of uncertain inputs and δ is the number of nodes from each of

these inputs. The weights (w_{k_j}), and sample locations (x_{k_j}) for the k^{th} point of j^{th} dimension are determined by implementing Gauss-Hermite, Gauss-Legendre, or Gauss-Laguerre quadratures based on the distribution of each uncertain input variable (Abramowitz et al., 1988). Then, the desired moments are calculated by Eq. 10.

$$\mu_i = E\left(g(X)^i\right) = \sum_{k_1=1}^{\delta} w_{k_1} \dots \sum_{k_n=1}^{\delta} w_{k_n} \times [g(x_{k_1}, \dots, x_{k_n})]^i \tag{10}$$

2.2.2. Univariate dimension reduction (UDR)

In the UDR method, the multivariate function, $g(X)$, is approximated to $\hat{g}(X)$ using the summation of several univariate functions (Rahman and Xu, 2004). Each of these univariate functions, $g_j(X_j)$, are similar to the original one, but each one is only a function of one variable with the remaining variables set to their mean values ($M_{\boldsymbol{X}_{j'}}$) (Eq. 11). Then, the original model output is approximated by the additive decomposition of the model (Eq. 12),

$$g_j(X_j) = g\left(X_j, X_{j'} = M_{\boldsymbol{X}_{j'}}\right) \quad \forall j, j' \in \{1,2,\dots,n\}, \qquad j \neq j' \tag{11}$$

$$g(X) \approx \hat{g}(X) = \sum_{j=1}^{n} g_j(X_{\boldsymbol{j}}) - (n-1)g(M_X) \tag{12}$$

where $g(M_X)$ is the model output value with all uncertain input variables set at their mean values. The moments of the output function ($g(X)$) are calculated using the estimated model outputs, $\hat{g}(X)$, and the quadrature formula (Eq. 13), similar to the FFNI method. Employing univariate quadrature formula with m nodes, the number of model evaluations is equal to $m = \delta n + 1$, since n univariate models are calculated at δ different nodes, and one extra model evaluation is performed with all variables set to their mean values.

$$\mu_i = E\left(g(X)^i\right) \approx E(\hat{g}(X)^i) = E\left(\left\{\sum_{j=1}^{n} g_j(x_j) - (n-1)\, g(M_x)\right\}^i\right) \tag{13}$$

2.2.3. *Sparse grid numerical integration*

Sparse Grid (SG) (Smolyak, 1963) is a numerical integration method that uses quadrature formulas, similar to FFNI, for estimating integrals (such as Eq. 9 and Eq. 10). The sample points and the weights are determined using Eq. 14 and Eq. 15, respectively (Xiong et al., 2010).

$$\vec{U}_n^k = \bigcup_{k+1 \le |i| \le k+n} U_1^{i_1} \otimes U_1^{i_2} \dots \otimes U_1^{i_n} \tag{14}$$

$$w_l = (-1)^{k+n-|i|} \binom{n-1}{k+n-|i|} \left(w_{j_1}^{i_1} \dots w_{j_n}^{i_n} \right) \tag{15}$$

where $\vec{U}_n^k$, is $P_s \times n$ array of all sample points with an accuracy of k, n denotes the number of dimensions, and P_s is the number of resulting possible sample points given $k+1 \le |i| \le k+d$. The operation $\otimes$ corresponds to the tensor product of arrays. The variables $U_1^{i_1}, \dots, U_1^{i_n}$ are one-dimensional quadrature points for each dimension, and i_α is the number of nodes in dimension α. Variable $|i|$ is the summation of the multi-indices ($|i| = i_1 + \cdots + i_n$). In Eq. 15, w_l is the weight for the l^{th} sample point $\overrightarrow{X_l} = \left[X_{j_1}^{i_1}, \dots, X_{j_n}^{i_n} \right] \in \vec{U}_n^k$, where $j_\alpha \in \{1, \dots, i_\alpha\}$. The parameter $w_{j_{i_1}}^{i_1}$ is the weight for the sample sets of one-dimensional quadrature. The i^{th} moment (μ_i) is calculated using Eq. 16.

$$\mu_i = E\left(g(X)^i\right) = \sum_{l=1}^{P_s} w_l \; g(\overrightarrow{X_l})^i \tag{16}$$

2.3. *Functional expansion–based methods*

In functional expansion-based methods, the model output is approximated by a polynomial function. This approximate simpler model is used in conjunction with UP methods to estimate the statistical moments. In this study, PCE is selected as a representative functional expansion–based method.

2.3.1. *Polynomial chaos expansion (PCE)*

Polynomial chaos expansion (Wiener, 1938) approximates the model output using orthogonal polynomials. It projects the output variable as a function of random variables with a specific distribution based on orthogonal stochastic polynomials (Anthony, 2013; Crestaux et al., 2009). The statistical moments

of the output are calculated using the projected polynomial expansion. The general form of the PCE of a random variable, $u(\theta)$, can be written as Eq. 17 (Crestaux et al., 2009).

$$u(\theta) = c_0\Gamma_0 + \sum_{i_1=1}^{\infty} c_{i_1}\,\Gamma_1\left(\xi_{i_1}(\theta)\right) + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1} c_{i_1 i_2}\Gamma_2\left(\xi_{i_1}(\theta),\xi_{i_2}(\theta)\right) + \cdots \tag{17}$$

In Eq. 17, Γ_p are the orthogonal polynomials of order p. Different orthogonal polynomials are used depending on the distribution of the random variable θ. For example, Hermite and Legendre are the polynomial basis for normal and uniform distributions, respectively. $\xi_i(\theta)$ is the standard variable, e.g., standard normal variable if the distribution for input(s) is normal, and c_i's are deterministic coefficients. Eq. 17 can be approximated using Eq. 18:

$$u(\theta) = u(\xi_1, \xi_2, \dots, \xi_n) \approx \sum_{j=0}^{\infty} b_j\psi_j(\xi(\theta)) \tag{18}$$

$$y = g(\xi) \approx \sum_{j=0}^{\frac{(p+n)!}{p!n!}-1} b_j\psi_j(\xi) \tag{19}$$

$$b_i = \frac{E[y\psi_j(\xi)]}{E[\psi_j^2(\xi)]} \tag{20}$$

where ψ_j is the j^{th} component of the orthogonal polynomials. In Eq. 18, b_j correspond to $c_{i_1 i_2 \dots i_p}$'s and are calculated based on Eq. 20. The output (y) for the models with multiple input variables, $g(\xi)$, can be approximated using n-dimensional PCE with an order of p (Eq. 19). For inputs not distributed normally, either specific polynomials are used, or the transformation of the variables to the standard normal variables is carried out.

3. Computational experiments

For computational experiments, all UP methods are implemented for propagating input uncertainty to the outputs of a set of test functions with known analytical forms. Numerous test functions and input distributions are considered in the experiments for studying the impacts of functional forms, the number of

uncertain inputs, and distributions on the performance of the UP methods. The test function names, their formulas, and the input distributions are summarized in the Supplementary Materials. The uncertainty propagation is carried out by calculating the first four statistical moments of the function outputs. The computational experiments start with three function evaluations and terminate at 1×10^6 function calls. At each increment, four moments are estimated using all applicable UP methods.

In the PCE method, the polynomials are truncated at the order of p. In this study, four different values of p, $p = \{2,3,4,5\}$, are considered to observe the impact of the polynomial truncation order. Also, we use three methods, Sobol sequence (PCE-S), Halton series (PCE-H), and FFNI (PCE-F), for estimating the numerator of Eq. 20.

The quality is defined as the minimum number of function evaluations required for a statistical moment to reach and remain within the desired error gap. To ensure that the estimated value stays within the 5% error gap, we started the search for the minimum number of function calls from the highest value then decreased until the estimation violated the 5% error gap of the true moment value. The error gap is a band with a width equal to a pre-determined error percentage of the '*true*' moment value. Four different error gaps, 2%, 5%, 10%, and 20%, were considered in the computational experiments. Most methods did not yield estimates within 2% error gap given the maximum budget. On the other hand, the trends for different UP methods were similar for 5%, 10%, and 20% error gaps; hence, the 5% error gap results are shown and analyzed in this paper. The '*true*' values of the moments are obtained using Monte Carlo simulation with 5×10^6 function evaluations. Experiments are implemented in Python 3.6. The packages Sobol_seq (Naught101, 2017) and Chaospy (Feinberg and Langtangen, 2015) are utilized for the Sobol series and PCE, respectively.

Four different cases are considered for assessing the performance of the UP methods. In each case, the effects of a specific factor are studied. The functions implemented in each case are included in Table S3 in Supplementary Materials.

3.1. Impact of nonlinearity

The performance of UP methods for various nonlinear functions is studied using two groups of functions. The first group contains twenty different nonlinear functions with one uniformly distributed input. The source of nonlinearity stems from exponential and trigonometric functions and the absolute value operator. The second group contains one-dimensional power functions, in which the input is uniformly distributed. The UP methods are evaluated for power functions with exponent values ranging from one to five to investigate the accuracy and efficiency of their estimates for increasing nonlinearity in a model.

3.2. Impact of the number of uncertain inputs

The effects of the number of uncertain inputs on the UP methods are studied by changing the dimensions of G (Surjanovic and Bingham, 2013) and Ackley (Surjanovic and Bingham, 2013) functions. It is assumed that all inputs are uncertain and uniformly distributed. The number of inputs is increased from one to five for the G function and to eleven for the Ackley function.

3.3. Impact of the uncertain input distribution

The impact of input distributions is investigated via two test sets. In the first set, UP methods are implemented and evaluated for different one-dimensional functions with both uniform and lognormal distributions. The second set includes the Ackley function with varying input dimensions (from one to eleven) and different distributions, uniform, normal, and lognormal.

3.4. General performance

As the final case, we consider all the 95 different test functions with various properties to assess the general performance of the UP methods. The results are studied to observe and deduct general trends for each UP method.

3.5. Application of the UP methods to Borehole and Steel Column models

The **Borehole** and **Steel Column** models (Surjanovic and Bingham, 2013), Eq. 21 and Eqs. 22-24, respectively, are used as problems to test the trends and the resulting recommendations generated using the

computational experiments. The **Borehole** model has eight input variables and calculates the flow rate in a borehole given its specifications. Table 1 lists the inputs with their distributions. The **Steel Column** model is nine-dimensional, and the input distributions are listed in Table 2. This model evaluates the reliability of a steel column based on the limit state function shown in Eq. 22, Eq. 23, and Eq. 24, which is a criterion of failure.

$$g(x) = \frac{2\pi T_u (H_u - H_l)}{\ln\left(r/r_w\right)\left(1 + \frac{2LT_u}{\ln\left(r/r_w\right) r_w^2 K_w}\right) + \frac{T_u}{T_l}} \tag{21}$$

$$g(x) = F_s - P\left[\frac{1}{2BD} + \frac{F_0 E_b}{BDH(E_b - P)}\right] \tag{22}$$

$$P = P_1 + P_2 + P_3 \tag{23}$$

$$E_b = \frac{\pi^2 EBDH^2}{2L^2} \tag{24}$$

Table 1. Distribution of uncertain inputs for Borehole function

Variable	Distribution
radius of borehole (m)	r_w ~ N (μ=0.10, σ=0.0161812)
radius of influence (m)	r ~ Lognormal (μ=7.71, σ=1.0056)
transmissivity of upper aquifer (m^2/y)	T_u ~ Uniform [63070, 115600]
potentiometric head of upper aquifer (m)	H_u ~ Uniform [990, 1110]
transmissivity of lower aquifer (m^2/yr)	T_l ~ Uniform [63.1, 116]
potentiometric head of lower aquifer (m)	H_l ~ Uniform [700, 820]
length of borehole (m)	L ~ Uniform [1120, 1680]
hydraulic conductivity of borehole (m/yr)	K_w ~ Uniform [9855, 12045]

Table 2. Distribution of uncertain inputs for Steel function

Variable	Distribution
yield stress (MPa)	F_s ~ Lognormal(mean=400, standard deviation=35)
deadweight load (N)	P_1 ~ N(μ=500000, σ=50000)
variable load (N)	P_2 ~ Gumbel(mean=600000, standard deviation=90000)
variable load (N)	P_3 ~ Gumbel(mean=600000, standard deviation=90000)

flange breadth (mm)	B ~ Lognormal(mean=300, standard deviation=3)
flange thickness (mm)	D ~ Lognormal(mean=20, standard deviation=2)
profile height (mm)	H ~ Lognormal(mean=300, standard deviation=5)
initial deflection (mm)	F_0 ~ N(μ=30, σ=10)
Young's modulus (MPa)	E ~ Weibull(mean=210000, standard deviation=4200)

4. Results and discussion

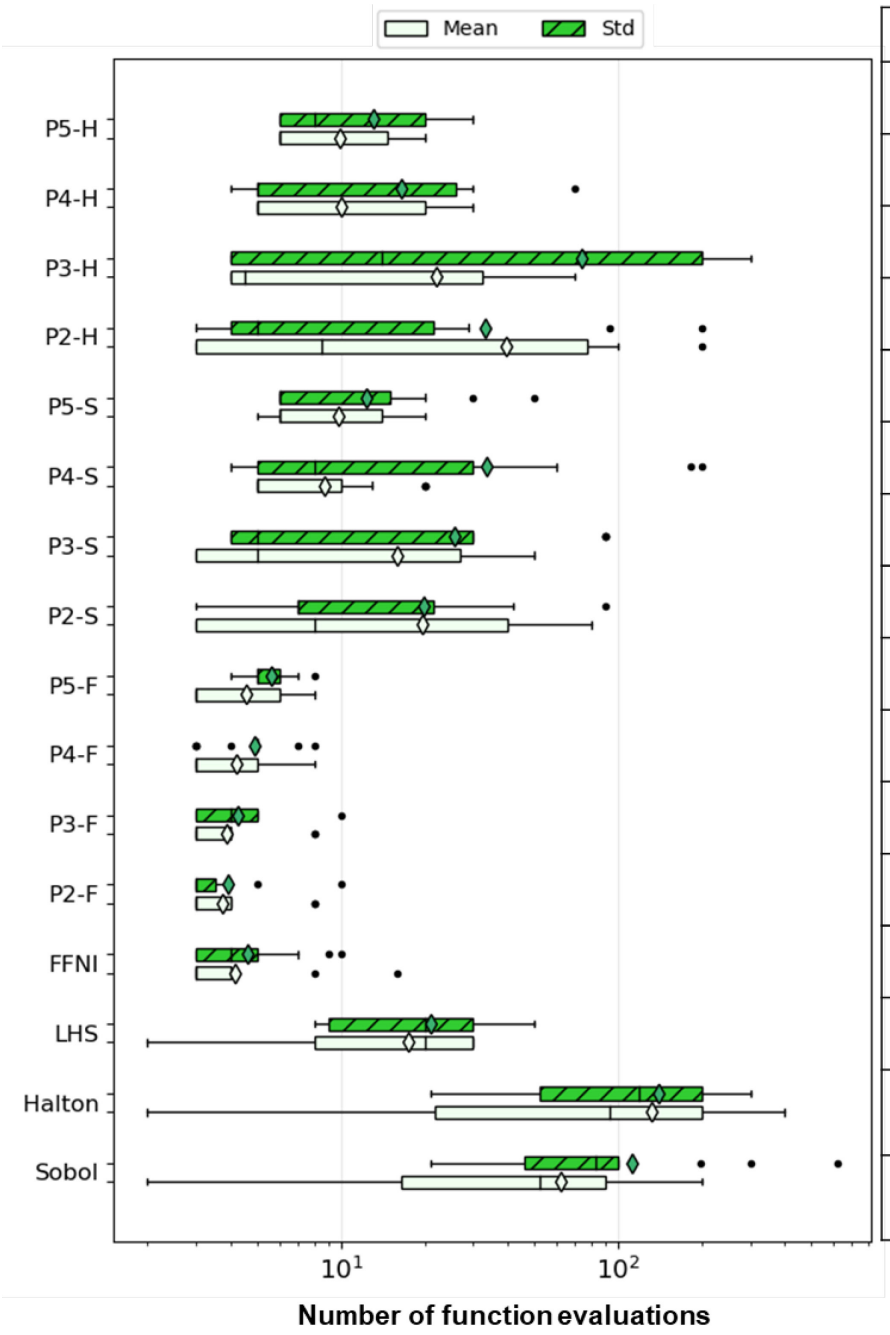
4.1. Impact of nonlinearity on the performance of uncertainty propagation methods

Figures 2 and 3 include boxplots for the minimum number of function evaluations required to converge to a 5% error gap for each of the first four moments for the first group of nonlinearity test functions. In the graphs, P(i)-F stands for the i^{th} order PCE where the integral was estimated using FFNI, P(i)-S using Sobol, and P(i)-H using Halton. The variables, n_1, n_2, n_3, and n_4, are the number of functions for which the method did not yield results within the 5% error gap of the *true* mean, standard deviation, skewness, and kurtosis values, respectively, with one million function evaluations. The plots do not depict results for UDR and SG because all test functions are one-dimensional, and for these functions, the UDR and SG revert to FFNI.

Figure 2 reveals that FFNI and P(i)-F required the lowest number of function evaluations, on average, to yield estimates that are within the 5% error envelope of the *true* mean and standard deviation. Their interquartile ranges and range of whisker values were the smallest for both mean and standard deviation estimations in comparison to all the methods (Figure 2). The outlier values for FFNI and P(i)-F were lower than the average outlier value of other UP methods. The use of MCS-based methods in PCEs resulted in PCE requiring, on average, a higher number of function evaluations than FFNI and P(i)-F to yield mean and standard deviation estimates within the 5% error gap. For PCE using Sobol (P(i)-S) and Halton (P(i)-H) sampling, the average number of function evaluations required and the interquartile range decreased with an increase in the PCE order. This behavior suggests that the higher-order polynomials aided in representing the nonlinearity in the test functions. However, PCEs with the lower orders, specially $p =$

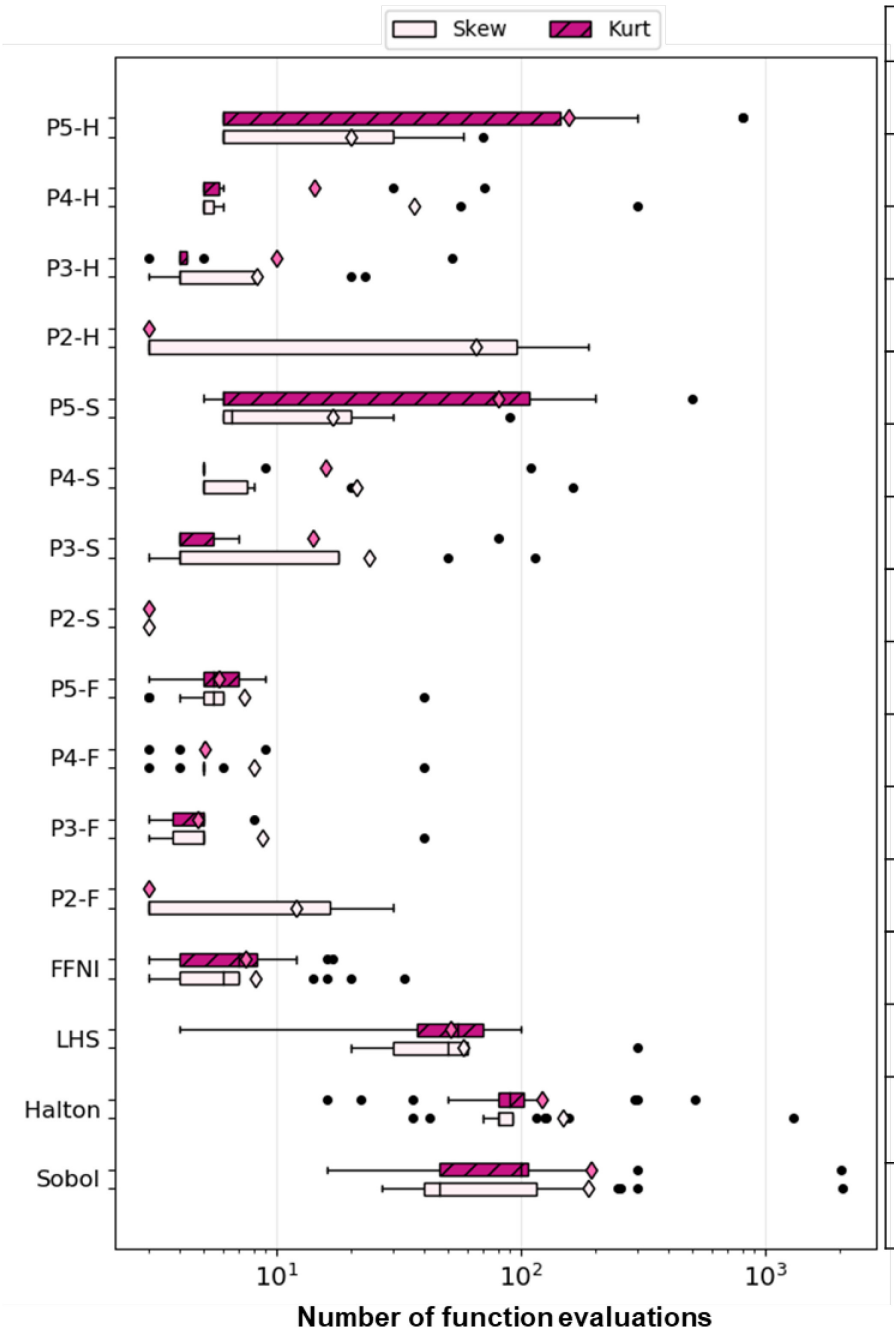
2, did not yield estimates within the 5% error gap of the *true* standard deviation for a large number of the test functions. The number of test functions, for which the standard deviation estimate was within the 5% error gap, increased as the order of the PCEs grew larger. The MCS-based methods, Sobol and Halton sampling, and LHS, needed more function evaluations to estimate the mean and standard deviation (Figure 2). Although the whiskers range was largest for MCS-based methods, the interquartile range was smaller than lower-order P(i)-S and P(i)-H suggesting a peaked distribution for the minimum number of function evaluations in comparison to the PCE methods with approximately same whisker range but larger interquartile ranges.

FFNI required the fewest number of function evaluations to estimate skewness and kurtosis (Figure 3) within the 5% error gap of the *true* values for all the functions. Both interquartile and whiskers ranges were smaller than other methods. Higher-order PCE where FFNI is used to calculate the integral (P(i)-F) estimated these two moments with a low number of function calls. However, they did not converge to a 5% error gap for all of the functions. The two lower-order P(i)-Fs did not converge to the error gap for more than 60% of the functions. Sampling using Sobol and Halton sequences and LHS yielded skewness and kurtosis estimates within the 5% error gap for all test functions; however, they required more function calls than FFNI and higher-order P(i)-F on average. The interquartile range of MC-based methods was significantly larger than that obtained by FFNI but smaller than what higher-order P(i)-S and P(i)-H methods yielded. The skewness and kurtosis estimates of PCEs were still not within the 5% error gap at the maximum allowed function evaluations for most functions, especially when lower-order PCEs were used (Figure 3). However, the number of functions where the estimates were not within 5% decreased as the order increased, suggesting that higher-order polynomials are necessary for estimating higher-order moments of the outputs for Sobol and Halton based PCEs. On the other hand, this is not true for PCEs based on FFNI, since FFNI is very efficient in estimating the moments for one-dimensional models, which leads to quick convergence to the desired error gap even with lower order polynomials. The number of outliers in estimating skewness and kurtosis is larger than the number observed when estimating the mean and standard deviation for all the methods.



Method	n_1	n_2
P5-H	0	1
P4-H	0	1
P3-H	0	3
P2-H	0	9
P5-S	0	1
P4-S	0	1
P3-S	0	3
P2-S	0	9
P5-F	0	1
P4-F	0	1
P3-F	0	3
P2-F	0	9
FFNI	0	0
LHS	0	0
Halton	0	0
Sobol	0	0

Figure 2. The minimum number of function evaluations for estimating mean, standard deviation within 5% of their *true* values for the test functions to study the impact of nonlinearity. n_1 and n_2 are the numbers of functions for which the method did not yield an estimate within a 5% gap of the *true* values for mean (Mean) and standard deviation (Std), respectively.



Method	n_3	n_4
P5-H	4	4
P4-H	9	10
P3-H	12	12
P2-H	17	18
P5-S	4	1
P4-S	9	10
P3-S	12	12
P2-S	18	18
P5-F	4	4
P4-F	9	8
P3-F	12	12
P2-F	17	18
FFNI	0	0
LHS	0	0
Halton	0	0
Sobol	0	0

Figure 3. The minimum number of function evaluations for estimating skewness and kurtosis within 5% of their *true* values for test functions to study the impact of nonlinearity. n3 and n4 are the numbers of functions for which the method did not yield an estimate within a 5% gap of the *true* values for skewness (Skew) and kurtosis (Kurt), respectively.

Figure 4 summarizes the results for the second group of test functions for evaluating the effect of nonlinearity, where different exponent values are used in the power function. Each figure demonstrates the minimum number of required function evaluations for each UP method to converge to a 5% error gap around the moment for Power functions with different exponent values. Based on Figure 4, MCS-based methods required a higher number of function evaluations to generate accurate estimates of both mean and standard deviation compared to the other methods. Furthermore, the number increased as the value of the exponent rose. The number of function evaluations required by FFNI and PCE where FFNI is used to estimate the integral (P(i)-F) did not change significantly for estimating the mean when the exponent value increased, and the change was the lowest for estimating standard deviation compared to the other methods. PCEs with second-order polynomials with integral estimates carried out using either Sobol or Halton sampling methods, P2-S and P2-H, needed more function calls than the higher-order PCEs for estimating mean and standard deviation of the output for power function with the large exponent values.

Monte Carlo simulation-based methods required more function evaluations to estimate skewness and kurtosis with increases in the exponent (Figure 4). The required function evaluations were the lowest, in general, for LHS for estimating skewness and kurtosis among the three sampling schemas. FFNI needed a significantly lower number of function evaluations than the MCS-based methods for accurate estimation of the skewness and kurtosis for the power function with all considered values of exponents. Furthermore, the change in the number of function calls as the power value rose was minimal compared to MCS-based methods. All PCEs with orders of two and three did not yield third- and fourth-moment estimates within the 5% error gap for the power function with an exponent of five due to high nonlinearity. As a result, higher orders of PCEs must be used when predicting higher moments.

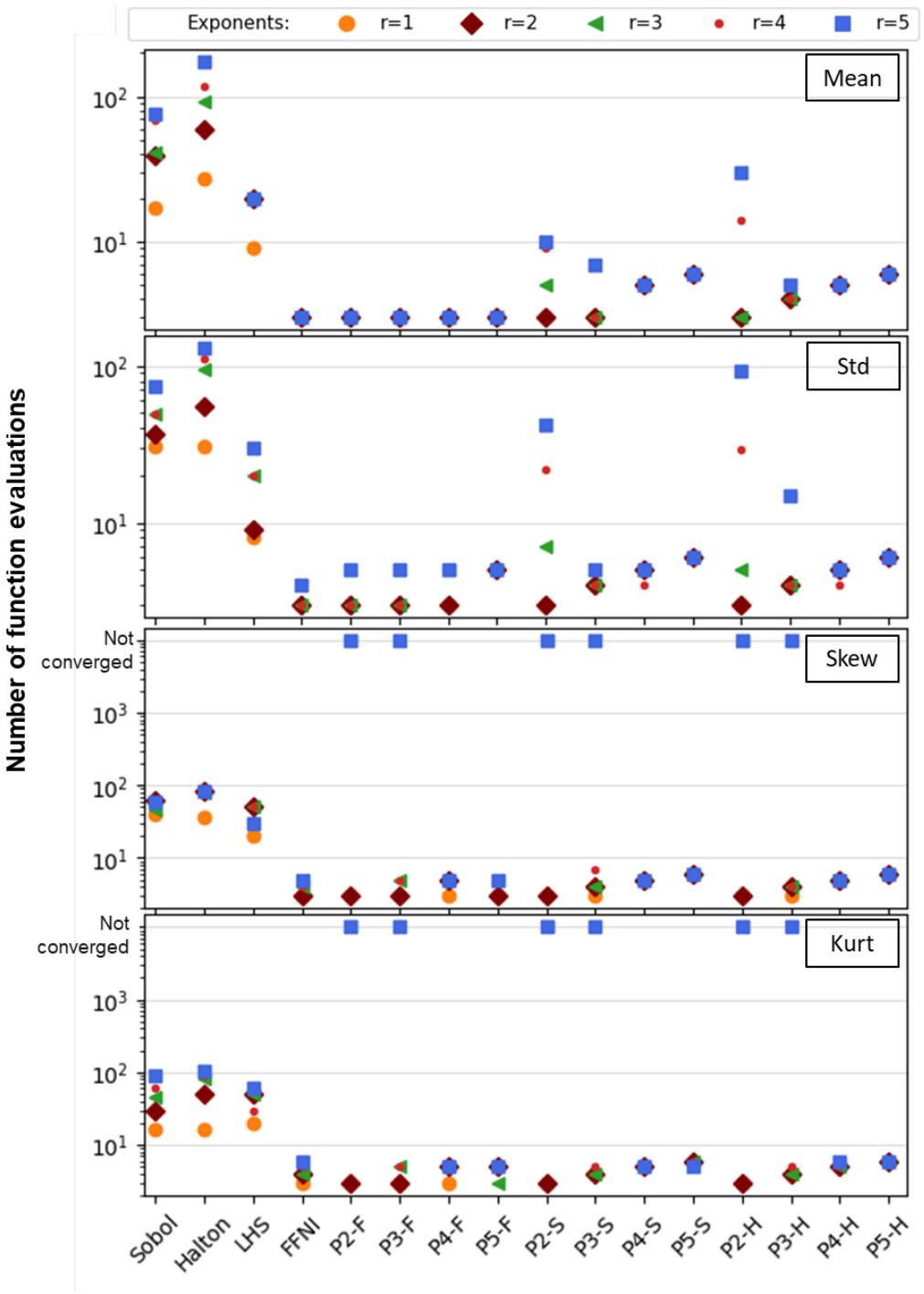


Figure 4. The minimum number of function evaluations for estimating mean (Mean), standard deviation (Std), skewness (Skew), and kurtosis (Kurt) within 5% of their *true* values for the power function with different exponents.

4.2. Impact of the number of uncertain inputs on the performance of uncertainty propagation methods

Figures 5 and 6 plot the minimum number of function evaluations needed by each method to achieve mean, standard deviation, skewness, and kurtosis estimates within the 5% error gap for G and Ackley functions. These plots show the impact of increasing the number of uncertain inputs on the performance of the UP methods. The plots reveal that the minimum number of required function calls to estimate the first four moments increases for all UP methods as the number of uncertain inputs increases. The increase is about an order of magnitude or more function evaluations for each input increment, and it is more significant for FFNI, SG, and P(i)-F for estimating all moments. This result is not surprising because the number of samples required increases exponentially with the number of uncertain inputs for these methods. The impact can be seen clearly for FFNI for the number of inputs above four in both functions. According to the plots (Figures 5 and 6), either the number of function calls to estimate all the moments using FFNI is above 10^5 or FFNI did not converge to the error gap for dimensions larger than four. However, SG converged to a 5% error gap for all moments of the G function and for the mean of the Ackley function for all dimensions within the maximum number of allowed function calls.

The UDR converged to the 5% error gap only up to five and three dimensions for the Ackley function (Figure 6) in estimating the mean and standard deviation, respectively, and did not yield accurate estimates of the other three moments for any of the dimensions. The results for high dimensional functions, especially estimation of higher moments, agree with the expected performance from UDR. The accuracy of the linear combination of univariate functions in representing the test function drop quickly for high dimensions, since a higher number of relations between the variables are overlooked by this method. In addition, higher moments have larger nonlinearity associated with them which is not well captured by univariate functions. Consequently, moment estimates obtained using these approximations did not converge to the error gap within the allowed number of function evaluations. The UDR is not applicable for G function (and hence is not included in Figure 5) because the univariate functions are equal to zero when G function is

approximated as the linear combination of the univariate functions (these formulas are given in Supplementary Materials).

Although the MCS-based methods required higher function evaluations, specially for lower dimensions, to estimate the moments, the number of function calls varied less with changes in the number of uncertain inputs compared to other methods. The results demonstrated the MCS-based methods as reliable approaches for estimating the moments as they yielded estimates of all four moments within the error envelope for all the functions with different dimensions. The PCEs where the integrals were approximated using Sobol and Halton sampling methods, in general, converged to the 5% gap of the *true* mean with a lower number of function evaluations compared to numerical integration methods and the PCEs associated with them. This is because the number of samples does not increase exponentially using the low-discrepancy series, unlike the quadrature-based methods. Furthermore, the rate of increase in required evaluations was slow as the dimension increased. The plots also reveal that, as a general trend, the number of required function evaluations increases as the polynomial order increases for PCEs (e.g., mean plot in Figure 6). According to Eq. 19, as the order and dimension of the polynomials for PCEs increase, the number of coefficients, b_i, increase, and based on Eq. 20, a larger number of samples provides more accurate estimates of these coefficients. The number of function evaluations required by PCEs is dependent on two factors, the order of the polynomial and the number of uncertain inputs of the model. Figure 6 illustrates that P2-S and P2-H did not yield estimates within the gap for any number of uncertain inputs for the Ackley function; however, the higher-order polynomials estimated the skewness and kurtosis within the 5% error gap, suggesting that lower-order polynomials were not representing the nonlinearity of functions with higher-order moments accurately. The MCS-based PCEs did not converge to the desired gap of the standard deviation for the G function with a higher number of uncertain inputs and almost none of the dimensions for estimating the skewness and kurtosis (Figure 5). We think the main reason for these results is the high nonlinearity in the G function due to several interaction terms of different uncertain inputs and absolute value function, and added nonlinearity for calculating the second to fourth moments of the output, which is making it difficult for the PCEs to capture the response surface of the G functions accurately.

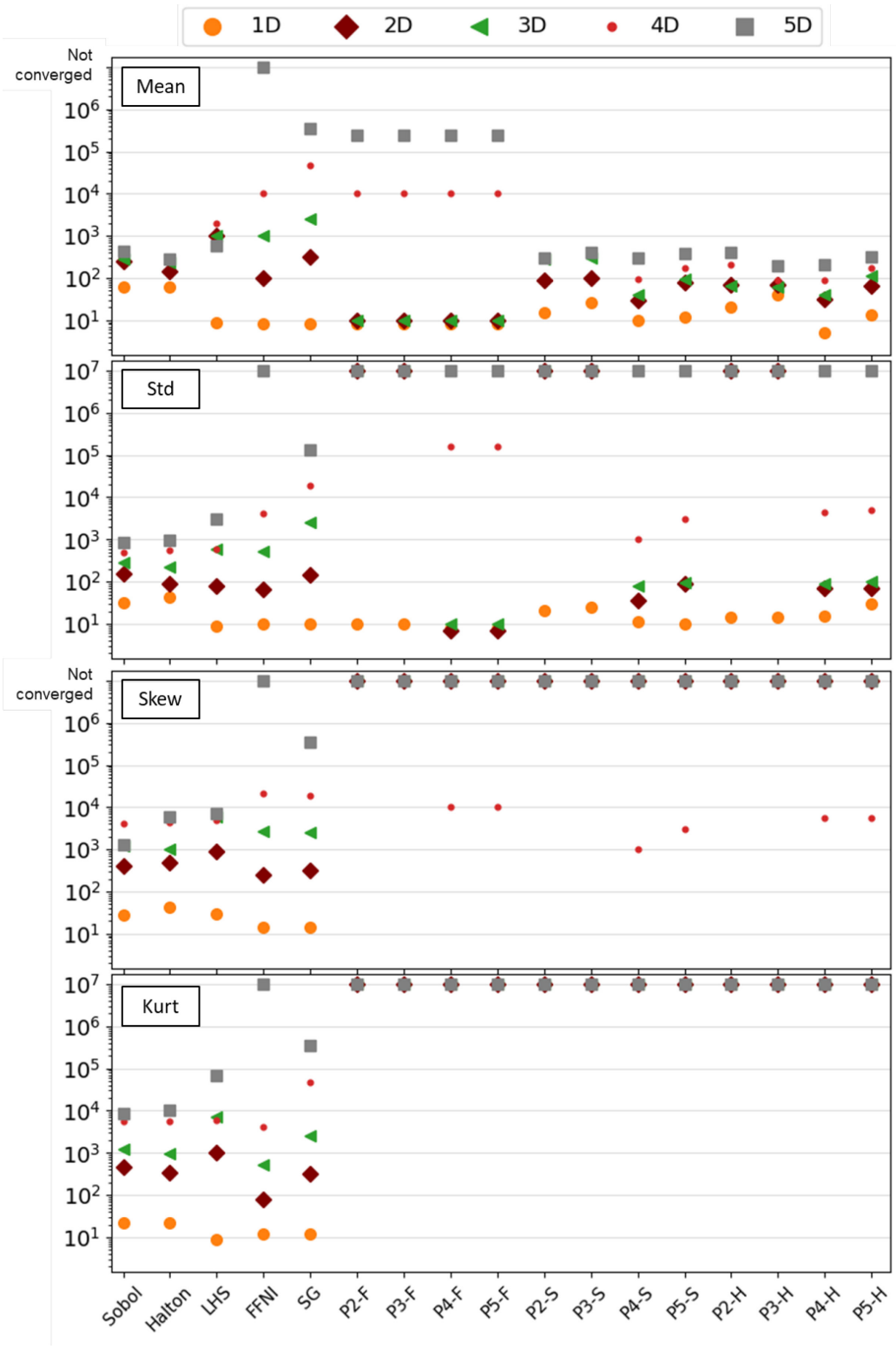


Figure 5. The minimum number of function evaluations for estimating mean (Mean), standard deviation (Std), skewness (Skew), and kurtosis (Kurt) within 5% of their *true* values for the case with the impact of dimensionality in G functions.

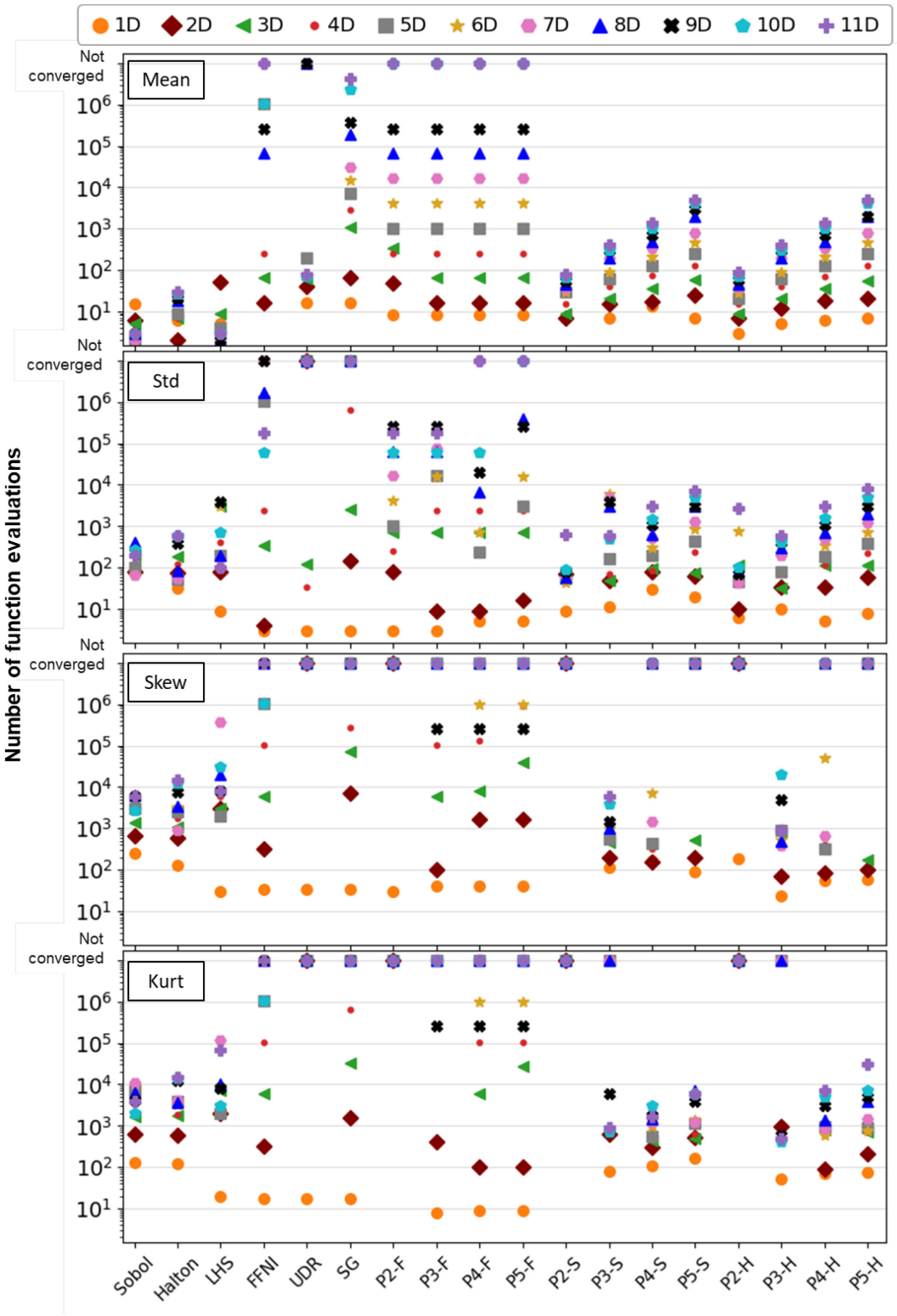


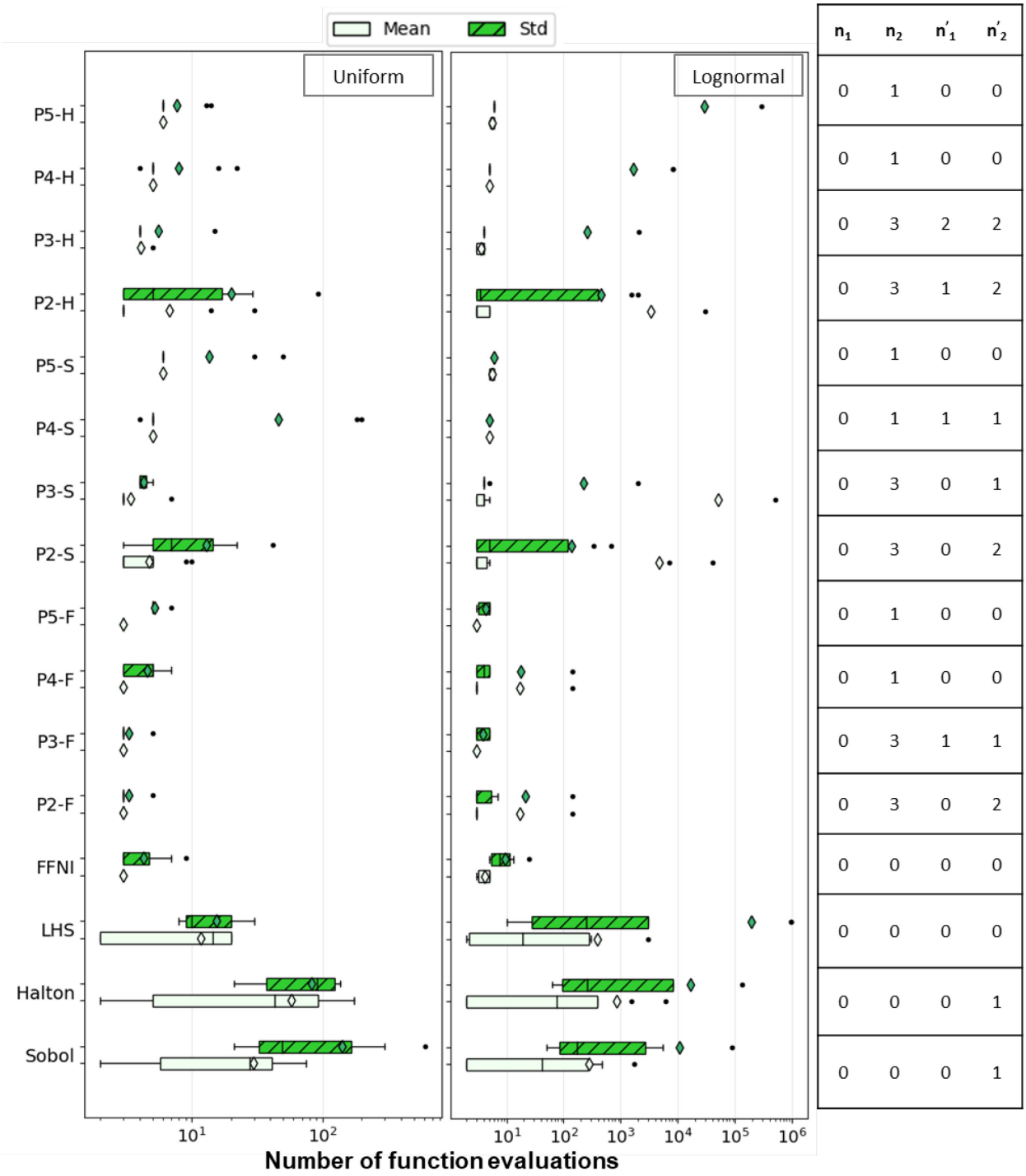
Figure 6. The minimum number of function evaluations for estimating mean (Mean), standard deviation (Std), skewness (Skew), and kurtosis (Kurt) within 5% of their *true* values for the case with the impact of dimensionality in Ackley functions.

4.3. Impact of the input distribution on the performance of uncertainty propagation methods

Figures 7 and 8 show the box plots of the minimum number of function evaluations to yield an accurate estimate of mean and standard deviation, and skewness and kurtosis, respectively, of the output for the one-dimensional test functions with input distributed uniformly and lognormally. The average number of function calls needed by each UP method is noticeably larger for the lognormal distribution than the uniform one, suggesting that input distribution is an important affecting factor. Additionally, the interquartile ranges increased for lognormal distribution in estimating all the moments with at least one order of magnitude for all methods, and the increase was the maximum for the MCS-based methods. However, the average was located far from the interquartile range for the lognormal distribution case, suggesting the effect of the outliers on the final value of the average and indicating that not all the functions require considerably higher function evaluations to converge to the error gap. Different nonlinear one-dimensional functions were used in the first test group, and similar to the results of the first case study group in Section 4.1, the MCS-based methods were the ones with the highest variability in results for both input distributions with the largest interquartile and whiskers ranges. Based on Figures 7 and 8, the MCS-based methods, on average, required the highest number of function evaluations in both uniform and lognormal distributions to converge to a 5% error gap of all moments, and the values were higher for the latter distribution.

The FFNI had the minimum median and average number of function calls to estimate all moments within the desired error gap for both input distributions. The change in average required function calls between two distributions was the lowest among all the methods. Different quadratures are used for selecting the nodes in FFNI, and the impact from the distributions is mitigated by the use of appropriate quadratures. The P(i)-F methods had comparable performance to the FFNI method in estimating the mean of the function outputs for both distributions. However, as the moment order increased, the number of function calls for the ones converged to the error gap and the number of functions for which the P(i)-F methods failed to converge to a 5% error gap for the last three moments increased (Figure 7 and 8). There

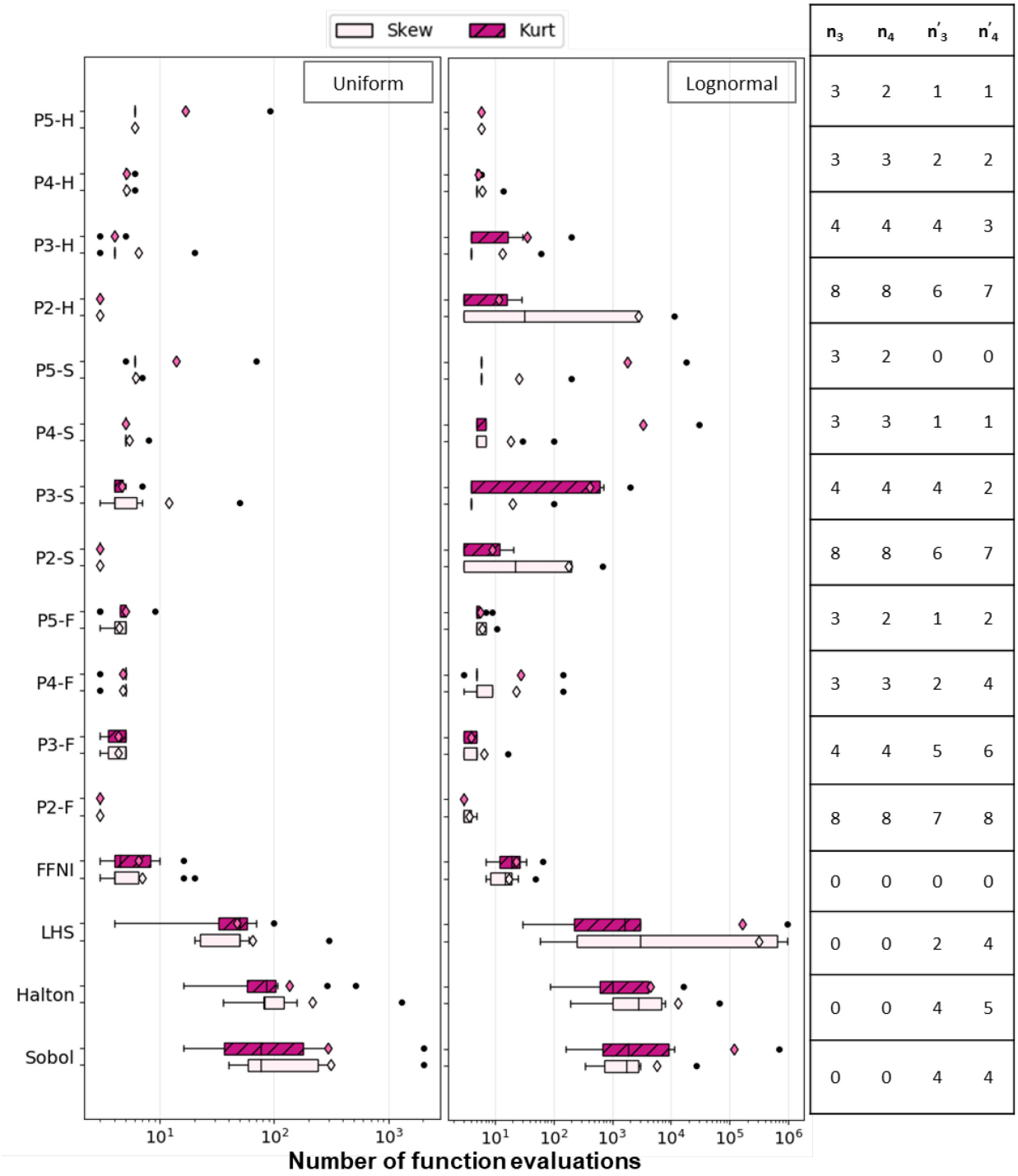
were no significant differences in the P(i)-F methods estimate quality between uniform or lognormal distributions, possibly due to the same reason for FFNI.



n_1	n_2	n'_1	n'_2
0	1	0	0
0	1	0	0
0	3	2	2
0	3	1	2
0	1	0	0
0	1	1	1
0	3	0	1
0	3	0	2
0	1	0	0
0	1	0	0
0	3	1	1
0	3	0	2
0	0	0	0
0	0	0	0
0	0	0	1
0	0	0	1

Figure 7. The minimum number of function evaluations for estimating mean (Mean) and standard deviation (Std) within 5% of their *true* values for case with impact of uniform distribution. n_1 and n_2 are the number of functions which did not converge to 5% gap of the *true* values for mean and standard deviation of uniform case study,

respectively. n_1' and n_2' are the number of functions which did not converge to 5% gap of the *true* values for mean and standard deviation of lognormal case study, respectively.



n_3	n_4	n'_3	n'_4
3	2	1	1
3	3	2	2
4	4	4	3
8	8	6	7
3	2	0	0
3	3	1	1
4	4	4	2
8	8	6	7
3	2	1	2
3	3	2	4
4	4	5	6
8	8	7	8
0	0	0	0
0	0	2	4
0	0	4	5
0	0	4	4

Figure 8. The minimum number of function evaluations for estimating skewness and kurtosis within 5% of their *true* values for case with impact of uniform distribution. n_3 and n_4 are the number of functions which did not converge to 5% gap of the *true* values for mean and standard deviation, respectively.

Lower order PCEs with integral approximated using MCS with Sobol and Halton sampling did not yield moments within the error gap of the *true* values for a large number of functions. Higher-order PCEs were better in estimating mean and skewness, where the plots (Figures 7 and 8) do not demonstrate a considerable fluctuation in the number of function evaluations needed by these methods between two different input distributions. The average number of function evaluations required increased significantly for lognormal distribution in estimating the standard deviation and kurtosis. The trends and performance of the UP methods for both distributions were in agreement with the ones observed in Section 4.1. suggesting that those results can be extended to different distributions.

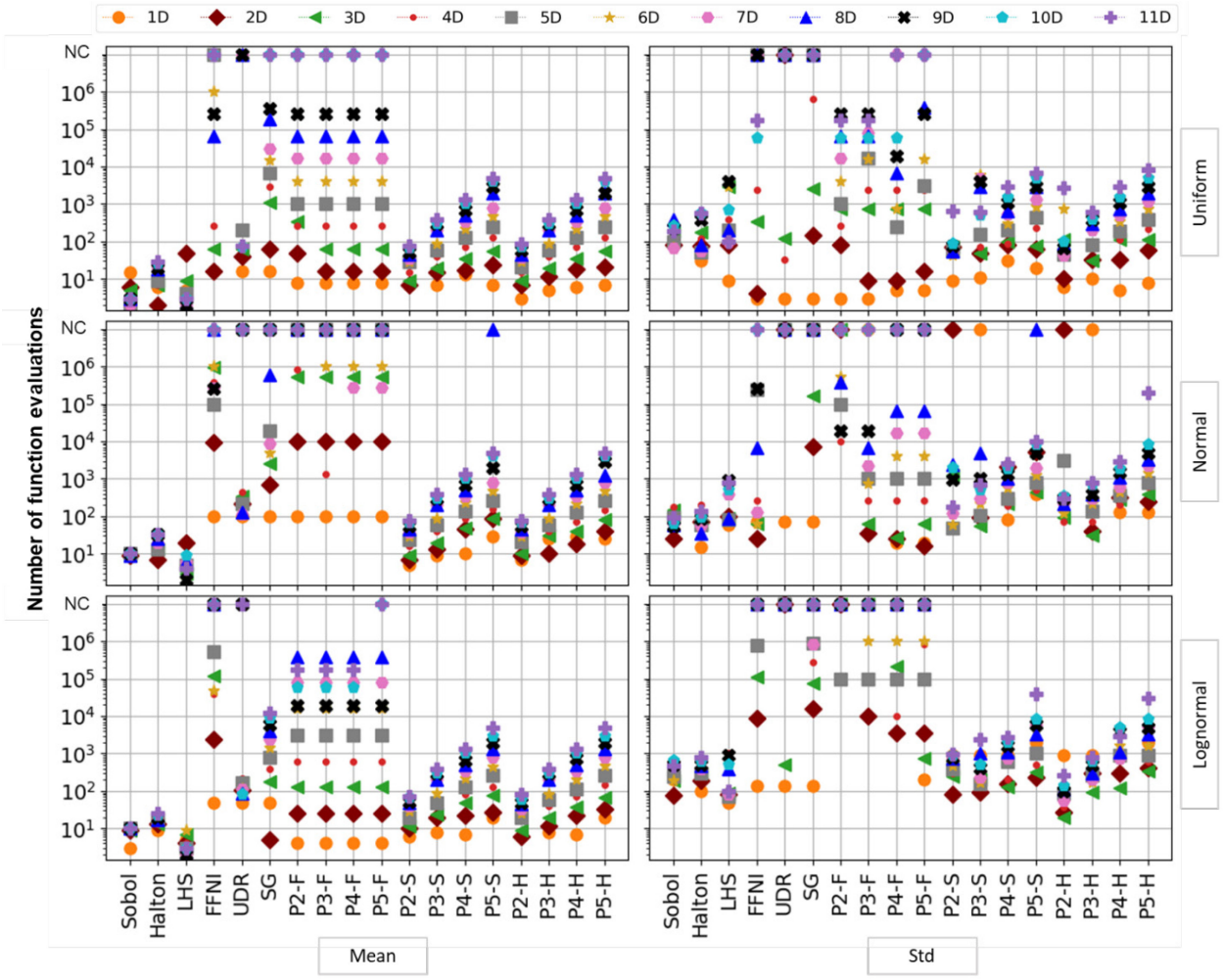


Figure 9. The minimum number of function evaluations for estimating mean (Mean), standard deviation (Std) within 5% of their *true* values for the case with the impact of distributions in Ackley functions. NC (Not Converged) indicates the methods which were not able to converge to the desired gap within 106 function evaluations.

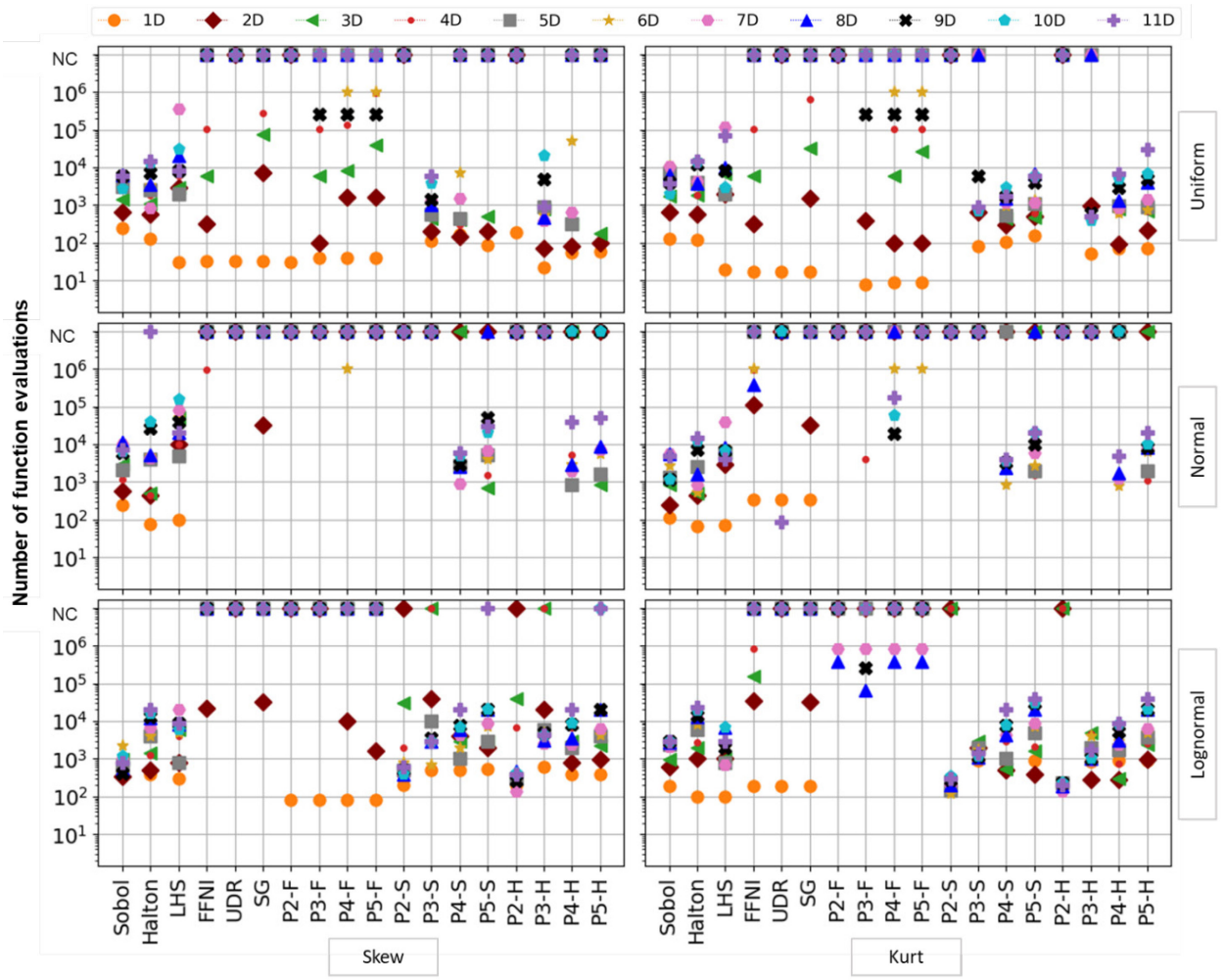


Figure 10. The minimum number of function evaluations for estimating skewness (Skew) and kurtosis (Kurt) within 5% of their *true* values for the case with the impact of distributions in Ackley functions. NC (Not Converged) indicates the methods which were not able to converge to the desired gap within 10^6 function evaluations.

In the second case study on the impact of input distributions, the minimum number of function calls to estimate the four moments for Ackley function with three different distributions for the inputs are shown in Figures 9 and 10. According to the figures, for the number of dimensions above 2, FFNI demonstrates large changes in the number of function evaluations to estimate the output moments among different distributions. The number of function calls does not change drastically for the UDR method in estimating the mean of the Ackley function. However, UDR was not able to converge to the error gap for most cases with higher moment orders. The performance of the MCS-based methods varies less, i.e., the required

number of function evaluations does not change significantly, in comparison to other methods. The PCEs based on Sobol and Halton sampling are not affected by the distribution of the input variables and require the same number of function evaluations for cases that converged to a 5% error gap of each moment.

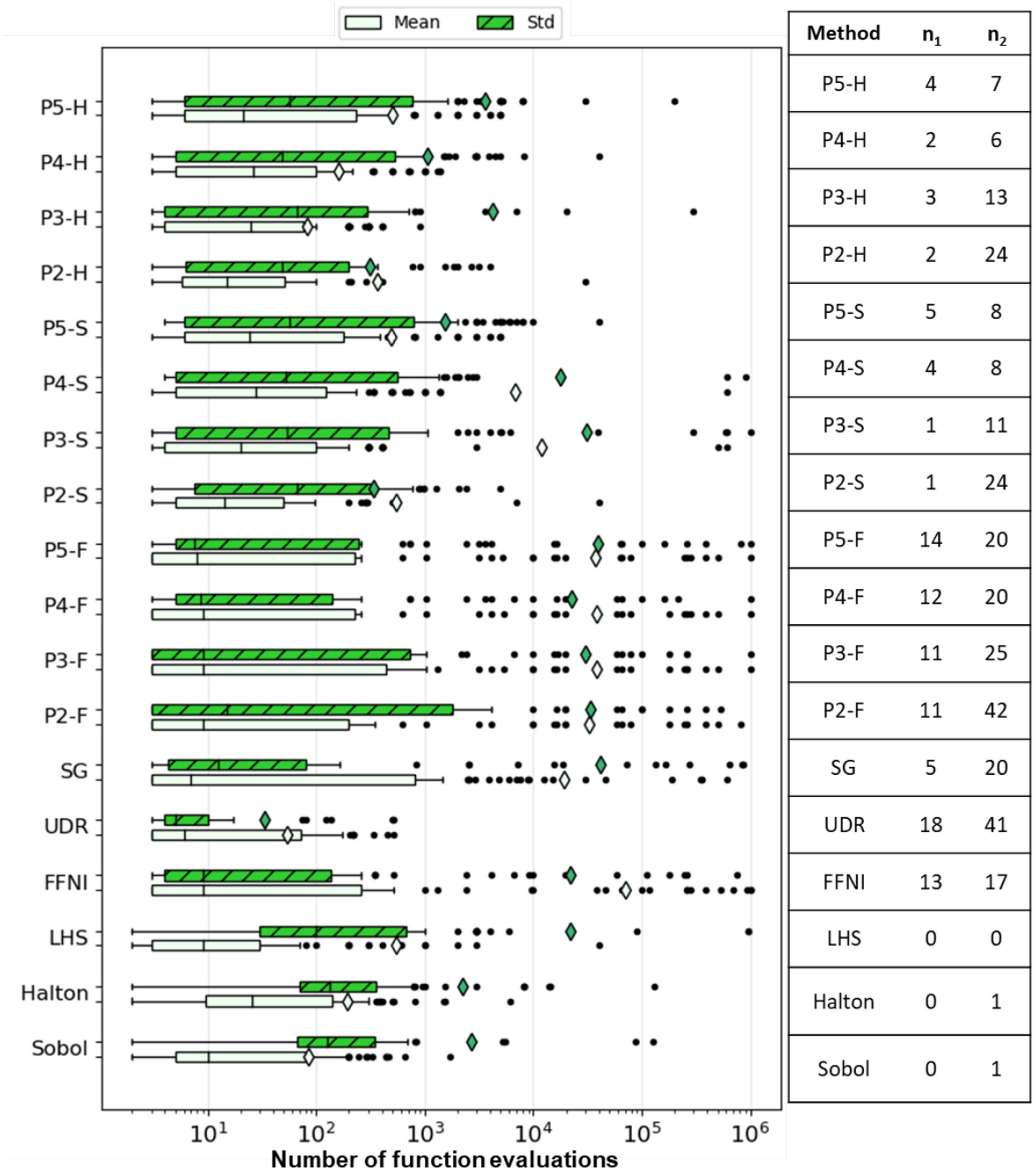
4.4. Comparison of the performance of uncertainty propagation methods for all test functions - Overall performance analysis

Figure 11 shows box plots of the minimum number of function evaluations required to estimate the mean and standard deviation within the 5% error gap of *true* moment values for all the test functions considered in this study. The results are important for problems where not all the characteristics of the models are known or given.

The plots in Figure 11 demonstrate that MSC-based methods outperformed the other methods in four aspects. First, they converged to the 5% error gap in estimating both mean and standard deviation for almost all the functions. Second, the median and the average number of required function evaluations are lower than the other UP methods. Third, the interquartile and whiskers ranges are the smallest in comparison to other methods converging to the error gap for the majority of the functions. Finally, the number of outliers is the lowest in comparison to any other method. The two prohibiting factors in estimating the moments, which are the dimensionality curse for numerical integration methods and the ability to represent the nonlinearity of the systems using PCEs, do not apply to MCS-based methods. Hence, they are more efficient in terms of the required number of function evaluations compared to all the other ones in this study.

Numerical integration methods and P(i)-F are deemed to be the least efficient UP methods as the minimum number of required function evaluations to estimate the mean and standard deviation using these methods is strongly affected by the number of uncertain inputs and grows quickly. There were many test functions with more than one uncertain input. As a result, FFNI, UDR, and P(i)-F were not able to yield accurate estimates of the first two moments for a larger number of the test functions, more than 10%, in comparison to other methods. However, among the numerical integration methods, SG converged to the error gap within the allowed number of function evaluations for a higher number of functions, and the

medians of the number of function calls were less than the medians of MCS-based methods. As the order of the moments increased, the number of functions for which the estimation of standard deviation through SG was unsuccessful grew larger.



Method	n_1	n_2
P5-H	4	7
P4-H	2	6
P3-H	3	13
P2-H	2	24
P5-S	5	8
P4-S	4	8
P3-S	1	11
P2-S	1	24
P5-F	14	20
P4-F	12	20
P3-F	11	25
P2-F	11	42
SG	5	20
UDR	18	41
FFNI	13	17
LHS	0	0
Halton	0	1
Sobol	0	1

Figure 11. The minimum number of function evaluations for estimating mean, standard deviation within 5% of their *true* values for the case of general performance. n_1 and n_2 are the numbers of functions that did not converge to a 5% gap of the *true* values for mean (Mean) and standard deviation (Std), respectively.

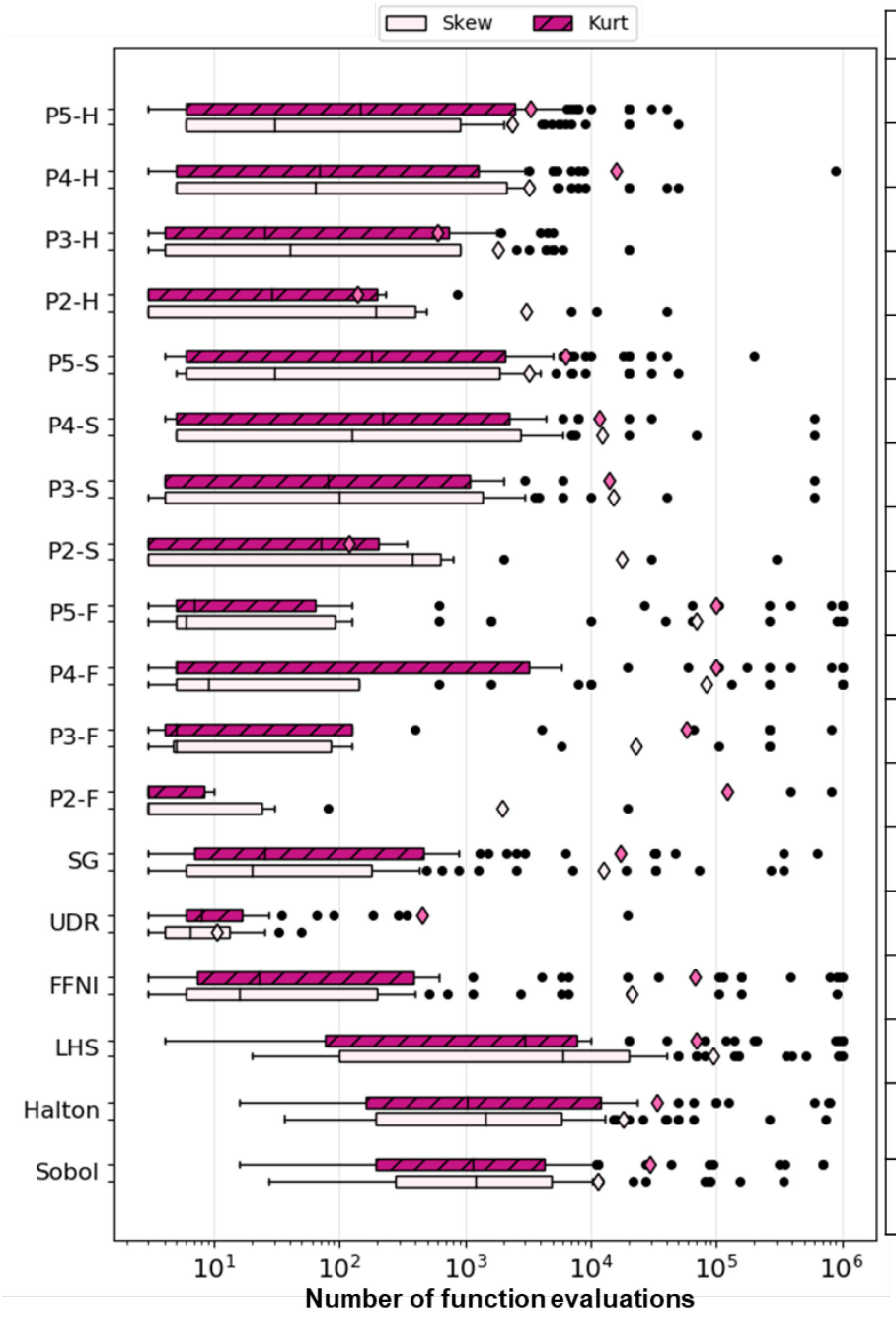
According to Figure 11, the function calls required by the P(i)-S and P(i)-H to converge to the desired error gap of mean does not have decreasing trend as the order of the polynomials increases, which we assume was due to the low order of the mean as the first moment where the degree of nonlinearity is lower than other moments. On the other hand, the change in the number of functions that these methods did not converge to the error gap of standard deviation, n_2, is notable as the order of the polynomial rises in value. The P5-H and P5-S managed to make estimations of the standard deviation within a 5% error gap for more than 90% of the test functions, which is larger than 75% of the functions, which were estimated using the second-order PCEs. The mean number of function calls used for the first two moment approximations by P(i)-S and P(i)-H were lower than numerical integration methods and P(i)-Fs, but it was significantly larger than MCS-based methods. PCEs do not guarantee convergence as the number of function calls increases, whereas, with MCS-based models, the estimation converges to the desired value at some point if the number of function calls is large enough. If the PCE with chosen order does not represent the nonlinearity accurately, then increasing the number of function evaluations would give accurate estimates of the built PCE, which is not necessarily representative of the desired function. On the other hand, MCS-based methods, using space-filling or low discrepancy sampling methods, cover larger amounts of the input(s) space as the number of samples and consequently the number of function evaluations increase and yield more accurate estimates of the moments as a result of the higher number of data.

Figure 12 illustrates the box plots of minimum required function calls to estimate the skewness and kurtosis within a 5% error bound of the *true* values for all the 95 test functions used in this paper. MCS-based methods estimated the third and fourth moments accurately for more than 90% of the test functions. Thus, they were the most reliable methods in terms of converging to the desired error gap. FFNI and SG predicted skewness and kurtosis of a larger number of functions, more than 65% and 75%, respectively, in comparison to the UDR and P(i)-F, which was less than 50% for estimating both skewness and kurtosis. The median of the number of required function calls for the functions the numerical integrations methods were able to converge to the desired error gap was lower than the median for MCS-based methods,

suggesting that if the numerical integrations are an appropriate method for the function characteristics, they are more likely to be efficient in estimating all moments accurately. Similar conclusions can be drawn for PCEs as well. For the cases in which PCEs converged to the 5% error gap of the *true* values of the moments, the mean and median of the number of the demanded function calls are less than the MCS-based methods. Hence, for the cases where the PCEs can approximate the nonlinearity of the model, the number of function evaluations is not large. Even though the performance of the PCEs was comparable to the MCS-based methods for estimating the mean, the performance significantly deteriorated for the higher moments, skewness and kurtosis, as can be seen with the higher values of n_3 and n_4 for PCEs versus MCS-based methods in Figure 12. The increase in the order of the polynomials improved the performance of the PCEs in converging to the error bound for a larger number of functions since the higher orders were able to represent the nonlinearity of the functions with better accuracy. The second-order PCEs had the highest number, more than 80%, of functions for which the convergence to the 5% error gap was not achieved. Therefore, based on Figure 12, they are not recommended to estimate the skewness and kurtosis.

4.5. Results for the application of the UP methods to Borehole and Steel Column models

Based on all the previous case studies, it is expected that MCS-based methods to be reliable methods and converge to the 5% error gap of all moments for both Steel and Borehole models and be one of the methods with a lower number of required model evaluations. Due to the high number of uncertain inputs for both models, SG among numerical integration methods is expected to have the best performance and require lower numbers of model calls if converged to the error gap. PCE-based methods are expected to have better or the same performance as MCS-based methods if they converge to the desired error gap of the *true* values of the moment, and the number of model evaluations should increase as the order of PCE gets larger.



Method	n_3	n_4
P5-H	31	20
P4-H	34	34
P3-H	49	50
P2-H	74	77
P5-S	28	20
P4-S	32	32
P3-S	49	49
P2-S	75	76
P5-F	46	47
P4-F	53	55
P3-F	66	65
P2-F	84	84
SG	31	27
UDR	54	48
FFNI	35	24
LHS	5	8
Halton	9	8
Sobol	10	9

Figure 12. The minimum number of function evaluations for estimating skewness and kurtosis within 5% of their *true* values for the case of general performance. n_3 and $n_4$are the numbers of functions that did not converge to a 5% gap of the *true* values for skewness (Skew) and kurtosis (Kurt), respectively.

The mean-standard deviation and skewness-kurtosis estimation for Steel and Borehole models are plotted in Figures 13 and 14, respectively. The figures show the minimum number of required model calls

used to converge to a 5% error gap of the moments. MCS-based methods were able to converge to the desired error gap of the mean with the lowest number of function evaluations. However, lower-order (2 and 3) PCEs that employed Sobol and Halton sampling also needed relatively low function evaluations to reach 5% of the *true* values. The number of function evaluations increased as the order increased, and for the Steel model, the PCE with orders of 4 and 5 did not converge to the chosen error gap, which suggests that these polynomials were not able to represent the nonlinearity of the Steel model. SG had the best performance among all the numerical integration methods with the lowest number of model evaluations to estimate all the moments within the 5% error gap of their *true* values. FFNI and the PCEs associated with it required a higher number of function calls due to the relatively high number of uncertain inputs in the models.

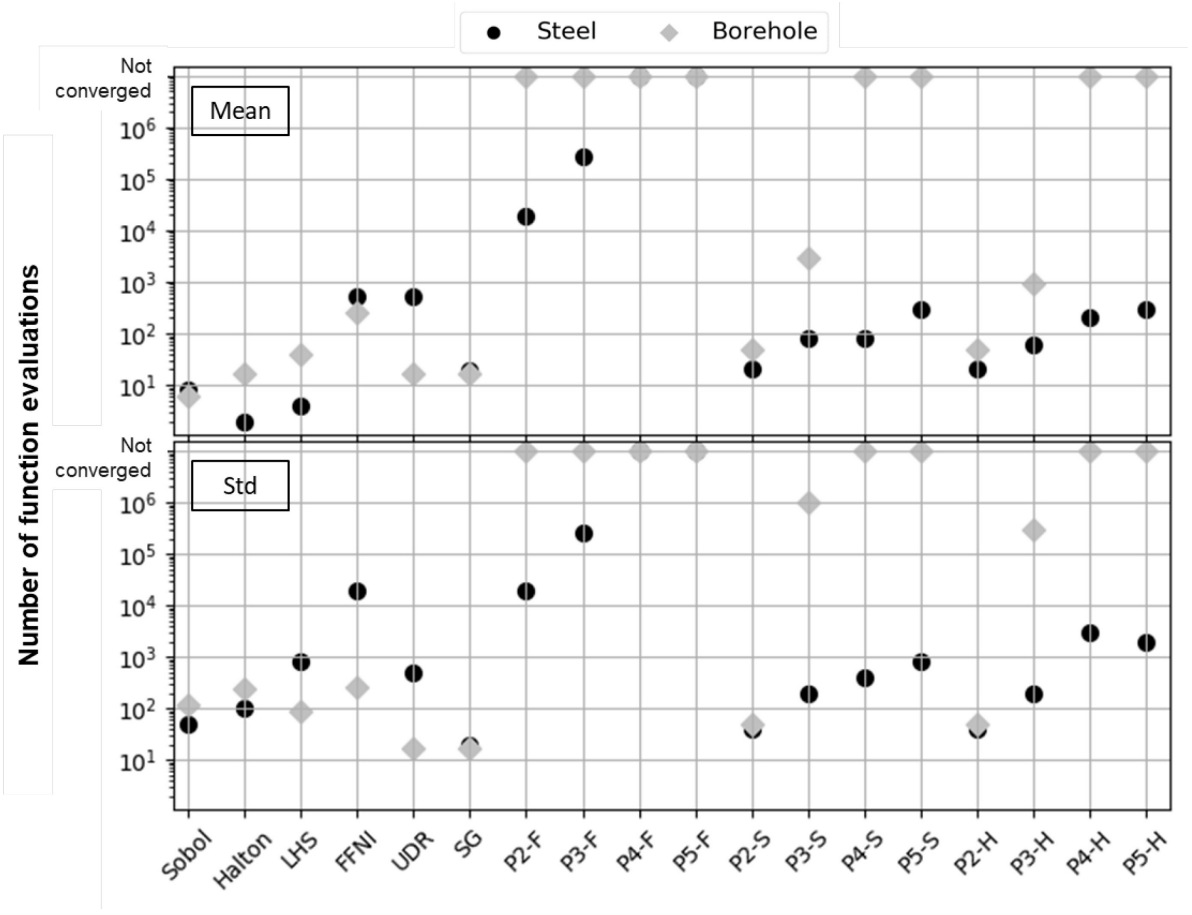


Figure 13. The minimum number of function evaluations for estimating mean, standard deviation within 5% of their *true* values for Steel and Borehole models.

For the Borehole model, none of the PCEs estimated the skewness and kurtosis within the desired error gap. The MCS-based methods and FFNI performed the best for this model, requiring the lowest number of function evaluations. On the other hand, PCEs converged to the desired gap with the number of function evaluations very close to the Sobol, Halton, and LHS for Steel model. There was an increasing trend for the required function evaluations of PCEs as the order got larger. The results for the Steel and Borehole models agree with the observations and conclusions drawn from the earlier case studies.

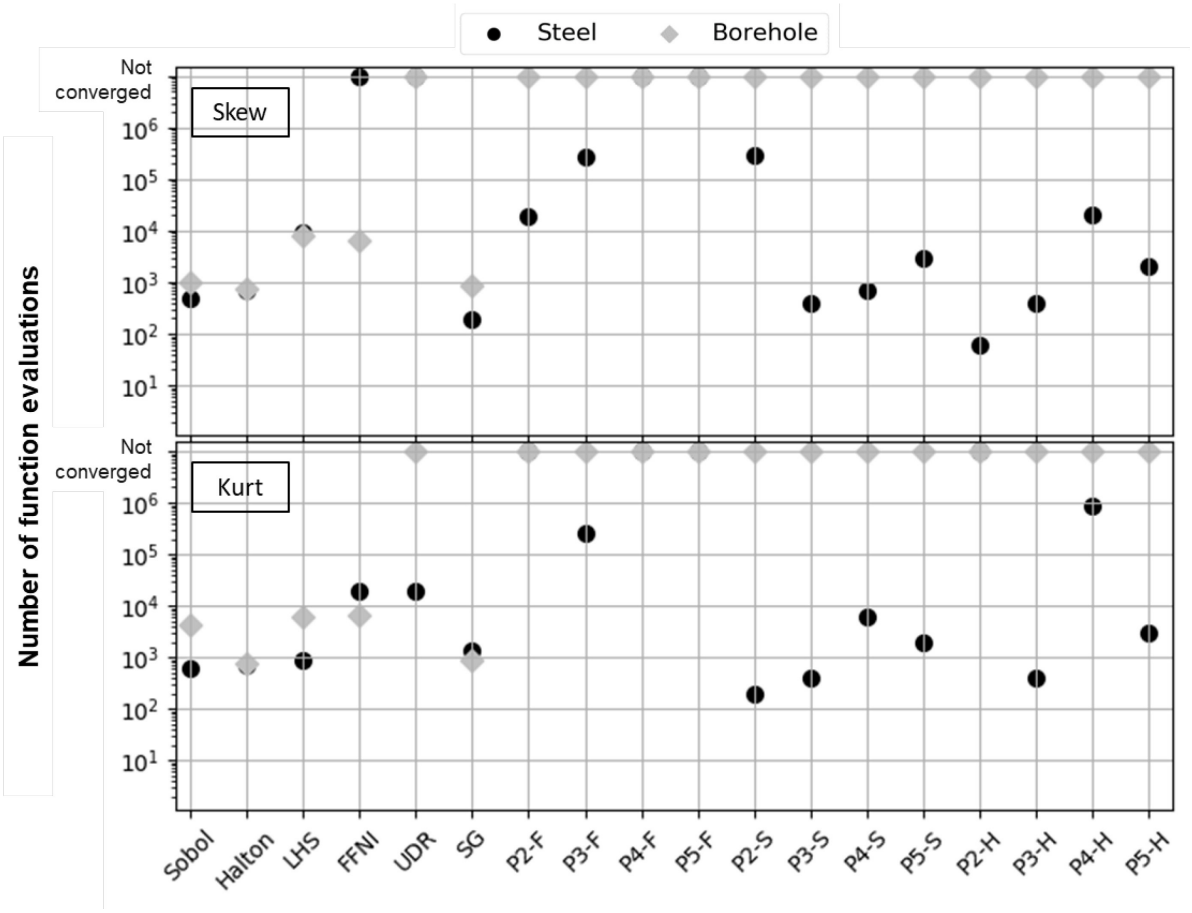


Figure 16. The minimum number of function evaluations for estimating skewness, kurtosis within 5% of their *true* values for Steel and Borehole models.

5. Conclusions and Future Directions

One important factor impacting the operation, design, and optimization of engineering processes is the uncertainties in the systems. The uncertainty propagation methods are used to quantify the uncertainty

of the system output, resulting from the uncertainty of inputs. This paper studied the performance of seven methods from three common groups of uncertainty propagation (UP) methods, including Monte Carlo simulation-based methods, numerical integration methods, and functional expansion-based methods, using computational experiments. The methods were Monte-Carlo method using Sobol series, Halton series, and Latin Hypercube sampling (LHS), numerical integration methods of Full Factorial Numerical Integration (FFNI), Univariate Dimension Reduction (UDR), and Sparse Grids (SG), and Polynomial Chaos Expansion (PCE) as the function expansion-based method. The study evaluated the impact of model characteristics, such as the number of uncertain inputs, non-linearity of the model, and uncertainty distribution type using 95 different test functions. The uncertainty propagation methods were compared based on the accuracy of the output estimates and the methods' efficiency in yielding these estimates. The accuracy was assessed using the first four statistical moments of the model output, and the efficiency was assessed using the minimum number of model calls required to reach and remain within the 5% error gap of the *true* values of these moments.

The efficiencies of the FFNI and PCEs that utilized FFNI had a strong dependence on the number of uncertain inputs. PCEs generally estimated the first two statistical moments accurately but did not converge to the desired error gap for skewness and kurtosis (third and fourth moments) for most test functions. The performance of Monte-Carlo simulation-based method was considerably impacted by the nonlinearity in the test functions compared to the other methods irrespective of the sampling technique used, and they required a higher number of model evaluations for complex functions. However, these methods were the most reliable UP methods and converged to the chosen error gap of all the moments for most (88%) test functions.

In light of our computational experiments, we constructed the following guidelines for selecting the UP method based on the characteristics of the model of interest: For models with less than five uncertain inputs regardless of the nonlinearity, FFNI is generally the most efficient method to estimate the first four moments. SG is a better choice among numerical integration methods, when the number of uncertain inputs is higher than five. As the number of uncertain inputs increases, for the models with high nonlinearities,

such as models that contain high power values, logarithmic, and trigonometric functions with possibly interacting terms, higher-order PCEs can be used to estimate the mean and standard deviation. However, higher-order PCEs are not the most reliable methods for estimating higher-order moments, such as skewness and kurtosis, because they did not converge to the error gap for most test functions. Although higher order PCEs may capture the nonlinearity of the model, it has to be noted that increase in order results in a higher number of coefficients that need to be estimated. This, in turn, translates into a lower efficiency for the PCE methods. There is a trade-off between the accuracy and efficiency with higher order PCEs. Finally, Monte Carlo simulation-based methods are recommended for models with high number of uncertain inputs and are reliable for estimating all four statistical moments of these models. The results from this study did not yield clear guidelines based on the distribution type as the trends observed were not consistent for the test functions.

Propagating the uncertainty of inputs and, potentially, parameters to stochastic high-fidelity simulation outputs is a potential area for utilizing the findings of this paper. In future work, we plan to select efficient uncertainty propagation methods using these guidelines for quantifying the uncertainty of stochastic surrogate models representing high fidelity stochastic simulation outputs.

6. Acknowledgment

This work was funded by NSF grant [1743445] and RAPID Manufacturing Institute, the U.S.A.

7. References

Abramowitz, M., Stegun, I.A., Romer, R.H., 1988. Handbook of mathematical functions with formulas, graphs, and mathematical tables.

Aleti, A., Trubiani, C., van Hoorn, A., Jamshidi, P., 2018. An efficient method for uncertainty propagation in robust software performance estimation. J. Syst. Softw. 138, 222–235. https://doi.org/10.1016/j.jss.2018.01.010

Allen, M.S., Camberos, J.A., 2009. Comparison of uncertainty propagation/response surface techniques for

two aeroelastic systems. Collect. Tech. Pap. - AIAA/ASME/ASCE/AHS/ASC Struct. Struct. Dyn. Mater. Conf. 1–19. https://doi.org/10.2514/6.2009-2269

Ankenman, B., Nelson, B.L., Staum, J., 2008. Stochastic kriging for simulation metamodeling. Proc. - Winter Simul. Conf. 362–370. https://doi.org/10.1109/WSC.2008.4736089

Anthony, O., 2013. Polynomial Chaos : A Tutorial and Critique from a Statistician's Perspective 1–20.

Burhenne, S., Jacob, D., Henze, G.P., 2011. Sampling based on sobol′ sequences for monte carlo techniques applied to building simulations. Proc. Build. Simul. 2011 12th Conf. Int. Build. Perform. Simul. Assoc. 1816–1823.

Crestaux, T., Le Maître, O., Martinez, J.M., 2009. Polynomial chaos expansion for sensitivity analysis. Reliab. Eng. Syst. Saf. 94, 1161–1172. https://doi.org/10.1016/j.ress.2008.10.008

Crombecq, K., Laermans, E., Dhaene, T., 2011. Efficient space-filling and non-collapsing sequential design strategies for simulation-based modeling. Eur. J. Oper. Res. 214, 683–696. https://doi.org/10.1016/j.ejor.2011.05.032

Duffy, J., Liu, S., Moskowitz, H., Plante, R., Preckel, P. V., 1998. Assessing multivariate process/product yield via discrete point approximation. IIE Trans. (Institute Ind. Eng. 30, 535–543. https://doi.org/10.1080/07408179808966493

Fahmi, I., Cremaschi, S., 2016. Computational Experiments on Sampling Methods for Uncertainty Propagation and the Implications for Simulation-Based Optimization, in: Computer Aided Chemical Engineering. Elsevier, pp. 1779–1784.

Gel, A., Garg, R., Tong, C., Shahnam, M., Guenther, C., 2013. Applying uncertainty quantification to multiphase flow computational fluid dynamics. Powder Technol. 242, 27–39. https://doi.org/10.1016/j.powtec.2013.01.045

Ghanem, R.G., Spanos, P.D., 1991. Spectral Stochastic Finite-Element Formulation for Reliability Analysis. J. Eng. Mech. 117, 2351–2372. https://doi.org/10.1061/(asce)0733-9399(1991)117:10(2351)

Grimmett, G., Stirzaker, D., 2001. Probability and random processes. Oxford university press.

Groen, E.A., Heijungs, R., Bokkers, E.A.M., de Boer, I.J.M., 2014. Methods for uncertainty propagation in life cycle assessment. Environ. Model. Softw. 62, 316–325. https://doi.org/10.1016/j.envsoft.2014.10.006

Halton, J.H., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numer. Math. 2, 84–90. https://doi.org/10.1007/BF01386213

Hansen, lars peter, 1982. Large Sample Properties of Generalized Method of Moments Estimators Author(s): Lars Peter Hansen Source: Econometrica 50, 1029–1054.

Hou, T., Nuyens, D., Roels, S., Janssen, H., 2019. Quasi-Monte Carlo based uncertainty analysis: Sampling efficiency and error estimation in engineering applications. Reliab. Eng. Syst. Saf. 191, 106549. https://doi.org/10.1016/j.ress.2019.106549

Hüllen, G., Zhai, J., Kim, S.H., Sinha, A., Realff, M.J., Boukouvala, F., 2019. Managing Uncertainty in Data-Driven Simulation-Based Optimization. Comput. Chem. Eng. 106519. https://doi.org/10.1016/j.compchemeng.2019.106519

Hunt, M., Haley, B., McLennan, M., Koslowski, M., Murthy, J., Strachan, A., 2015. PUQ: A code for non-intrusive uncertainty propagation in computer simulations. Comput. Phys. Commun. 194, 97–107. https://doi.org/10.1016/j.cpc.2015.04.011

Jia, X.Y., Jiang, C., Fu, C.M., Ni, B.Y., Wang, C.S., Ping, M.H., 2019. Uncertainty propagation analysis by an extended sparse grid technique. Front. Mech. Eng. 14, 33–46. https://doi.org/10.1007/s11465-018-0514-x

Joe, S., Kuo, F.Y., 2008. Notes on generating Sobol sequences Gray code implementation. English 2–4.

Klavetter, K., Posluszny, D., Warr, J., Cremaschi, S., Sarica, C., Subramani, H.J., 2012. Uncertainty analysis of multiphase flow models: A comparison of three propagation approaches. BHR Gr. - 8th North Am. Conf. Multiph. Technol. 259–271.

Lee, S.H., Chen, W., 2009. A comparative study of uncertainty propagation methods for black-box-type problems. Struct. Multidiscip. Optim. 37, 239–253. https://doi.org/10.1007/s00158-008-0234-7

Liu, Y., Gupta, H. V., 2007. Uncertainty in hydrologic modeling: Toward an integrated data assimilation

framework. Water Resour. Res. 43, 1–18. https://doi.org/10.1029/2006WR005756

Luo, Y. zhong, Yang, Z., 2017. A review of uncertainty propagation in orbital mechanics. Prog. Aerosp. Sci. 89, 23–39. https://doi.org/10.1016/j.paerosci.2016.12.002

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.

Miller, D.C., Ng, B., Eslick, J., Tong, C., Chen, Y., 2014. Advanced Computational Tools for Optimization and Uncertainty Quantification of Carbon Capture Processes, in: Eden, M.R., Siirola, J.D., Towler, G.P. (Eds.), Proceedings of the 8th International Conference on Foundations of Computer-Aided Process Design, Computer Aided Chemical Engineering. Elsevier, pp. 202–211. https://doi.org/https://doi.org/10.1016/B978-0-444-63433-7.50021-3

Murcia, J.P., Réthoré, P.E., Dimitrov, N., Natarajan, A., Sørensen, J.D., Graf, P., Kim, T., 2018. Uncertainty propagation through an aeroelastic wind turbine model using polynomial surrogates. Renew. Energy 119, 910–922. https://doi.org/10.1016/j.renene.2017.07.070

Padulo, M., Campobasso, M.S., Guenov, M.D., 2007. Comparative analysis of uncertainty propagation methods for robust Engineering Design. Proc. ICED 2007, 16th Int. Conf. Eng. Des. DS 42, 1–12.

Rahman, S., Xu, H., 2004. A univariate dimension-reduction method for multi-dimensional integration in stochastic mechanics. Probabilistic Eng. Mech. 19, 393–408. https://doi.org/10.1016/j.probengmech.2004.04.003

Rajabi, M.M., 2019. Review and comparison of two meta-model-based uncertainty propagation analysis methods in groundwater applications: polynomial chaos expansion and Gaussian process emulation. Stoch. Environ. Res. Risk Assess. 33, 607–631. https://doi.org/10.1007/s00477-018-1637-7

Safta, C., Chen, R.L.Y., Najm, H.N., Pinar, A., Watson, J.P., 2017. Efficient Uncertainty Quantification in Stochastic Economic Dispatch. IEEE Trans. Power Syst. 32, 2535–2546. https://doi.org/10.1109/TPWRS.2016.2615334

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput.

Phys. Commun. 181, 259–270. https://doi.org/10.1016/j.cpc.2009.09.018

Smolyak, S.A., 1963. Quadrature and interpolation formulas for tensor products of certain classes of functions, in: Doklady Akademii Nauk. pp. 1042–1045.

Sobol', I.M., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. Zhurnal Vychislitel'noi Mat. i Mat. Fiz. 7, 784–802.

Sofi, A., Muscolino, G., Giunta, F., 2020. Propagation of uncertain structural properties described by imprecise Probability Density Functions via response surface method. Probabilistic Eng. Mech. 60, 103020. https://doi.org/10.1016/j.probengmech.2020.103020

Surjanovic, S., Bingham, D., 2013. Virtual Library of Simulation Experiments: Test Functions and Datasets.

Tardioli, C., Kubicek, M., Vasile, M., Minisci, E., Riccardi, A., 2016. Comparison of non-intrusive approaches to uncertainty propagation in orbital mechanics. Adv. Astronaut. Sci. 156, 3979–3992.

Tripathy, R., Bilionis, I., Gonzalez, M., 2016. Gaussian processes with built-in dimensionality reduction: Applications to high-dimensional uncertainty propagation. J. Comput. Phys. 321, 191–223. https://doi.org/10.1016/j.jcp.2016.05.039

Wang, H., Sheen, D.A., 2015. Combustion kinetic model uncertainty quantification, propagation and minimization. Prog. Energy Combust. Sci. 47, 1–31. https://doi.org/10.1016/j.pecs.2014.10.002

Wiener, N., 1938. The homogeneous chaos. Am. J. Math. 60, 897–936.

Wong, T.-T., Luk, W.-S., Heng, P.-A., 2005. Sampling with Hammersley and Halton Points. Graph. Tools--The jgt Ed. Choice 255–270. https://doi.org/10.1201/b10628-32

Xiong, F., Greene, S., Chen, W., Xiong, Y., Yang, S., 2010. A new sparse grid based method for uncertainty propagation. Struct. Multidiscip. Optim. 41, 335–349. https://doi.org/10.1007/s00158-009-0441-x

Yang, S., Xiong, F., Wang, F., 2017. Polynomial Chaos Expansion for Probabilistic Uncertainty Propagation. Uncertain. Quantif. Model Calibration. https://doi.org/10.5772/intechopen.68484

Appendix: