two-stage approach of multiplicative dimensional reduction and polynomial chaos for global sensitivity analysis and uncertainty quantification with a large number of process uncertainties

Le Quang Minh[1a], Pham Luu Trung Duong[2a†], Jorge Goncalves[2,4], Ezra Kwok [3], Moonyong Lee[1†]

[1]School of Chemical Engineering, Yeungnam University, Gyeongsan, Rep. of Korea

[2]Luxembourg Centre for Systems Biomedicine, Systems Control Group, University of Luxembourg, Luxembourg

[3]Chemical & Biological Engineering, University of British Columbia, Vancouver, Canada

[4]Department of Engineering, University of Cambridge, Cambridge, UK

[a] These authors contributed equally to this work.

[†]Correspondence concerning this article should be addressed to:

Prof. Moonyong Lee

Email: *mynlee@yu.ac.kr*

Telephone: +82-53-810-2512

Dr. Duong Pham

Email: *pluutrungduong@yahoo.com*

Telephone: +352 4666449649

ABSTRACT

Uncertainties associated with estimates of model parameters are inevitable when simulating and modeling chemical processes and significantly affect safety, consistency, and decision making. Quantifying those uncertainties is essential for emulating the actual system behaviors because they can change the management recommendations that are drawn from the model. The use of conventional approaches for uncertainty quantification (e.g., Monte-Carlo and standard polynomial chaos methods) is computationally expensive for complex systems with a large/moderate number of uncertainties. This paper develops a two-stage approach to quantify the uncertainty of complex chemical processes with a moderate/large number of uncertainties (greater than 5). The first stage applies a multiplicative dimensional reduction method to approximate the variance-based global sensitivity measures (Sobol's method), and to simplify the model for the uncertainty quantification stage. The second stage uses the generalized polynomial chaos approach to quantify uncertainty of the simplified model from the first stage. A rigorous simulation illustrates the proposed approach using an interface between MATLAB and HYSYS for three complex chemical processes. The proposed method was compared with conventional approaches, such as the Quasi Monte-Carlo sampling-based method and standard polynomial chaos-based method. The results revealed the clear advantage of the proposed approach in terms of the computational efforts.

Keywords: Uncertainty quantification; Process uncertainty; Sensitivity analysis; Multiplicative dimensional reduction method, Polynomial chaos.

1. Introduction

Uncertainties associated with process variables are inevitable when modeling and designing chemical processes and can significantly affect safety, consistency, and decision making. Conventional process design based on a nominal case without considering uncertainties can have negative influences on the design accuracy. The problems of process design under uncertainties has attracted considerable attention recently especially regarding safety, reliability, and economic decisions[1]. At the design level, the uncertainties, which depend on several input parameters, are classified into two common types, including the epistemic and stochastic uncertainties[2]. Regarding the definition of input variables, propagation via the process of all these input uncertainties onto the process output of interest is one of the main tasks of uncertainty analysis[3].

Probabilistic approaches, such as Monte-Carlo (MC) and Quasi Monte-Carlo (QMC) methods, provide a common framework for the uncertainty quantification (UQ) and uncertainty propagation (UP) in the model input to its output[4-7]. MC/QMC methods generate an ensemble of random realizations from its uncertainty distribution to evaluate the model for each element of a sample set and estimate the relevant statistical properties, such as the mean, standard deviation, and quantile of output[8]. Furthermore, it can examine the different parameter values one by one and combinations using a more comprehensive approach, performing a global sensitivity analysis[9]. Sensitivity analysis (SA) of a process model aims to characterize how the process model outputs respond to the variation in inputs with an emphasis on finding the input parameters to which the outputs are the most sensitive [10,11]. Note that it is suitable for factor interactions and a non-linear relationships between factors and the output[12] because it is interested in the entire field of potential alterations of the input parameters.

Despite the simplicity in their implementation, however, the mean convergence was estimated to be in the order of $O(1/\sqrt{M})$, where M is the number of samples, which makes MC -based approaches computationally expensive and they are only used as a last resort. To tackle practical and time-consuming problems, Celse et al.[13] constructed an accurate, efficient-to-evaluate surrogate model that can be used in place of expensive simulations. Currently, there is growing demand for computationally efficient surrogate

models[14] that can ensure an acceptable degree of accuracy[15]. Similarly, quantifying the dependence on the uncertain parameters using a surrogate model for generalized polynomial chaos (gPC) expansion achieved faster convergence rate in various areas, such as modeling, control, robust optimal design, and fault detection problems[8]. The gPC method, which was first proposed by Wiener [16], is a spectral expansion of a random process based on the orthonormal polynomials in terms of the random variables. Nagy and Braatz [17] considered a polynomial chaos expansion for UQ and the robust design of a batch crystallization process. They reported that the gPC approach can reduce the computational cost for a system with a relatively small number of design inputs and uncertain parameters compared to MC/QMC methods. Shen and Braatz [18] developed a new polynomial chaos based algorithm on the design of batch and continuous-flow chemical reactors with probability uncertainties. Duong and Lee [19,20] examined a PID controller design using the gPC method. Du et al. [21] demonstrated a fault detection solution by combining the gPC with a maximum likelihood framework. Recently, Duong et al.[8] studied the problem of UQ /SA of chemical processes using the standard gPC method for systems with a small number of random inputs. Xiu and Karniadakis [22] proposed utilizing the Askey scheme for gPC expansion with non-standard distributions.

When smoothness in the output is approximated, the gPC expansion for engineering purposes with a uniform and Gaussian distribution exhibited rapid convergence; in some cases, even exponential convergence can be obtained [23]. A current limitation of the standard full gPC approach, where the coefficients are estimated using the tensor cubature, is that the number of model evaluations grows exponentially and may not be applicable to systems with a moderate/large number of uncertainties. In this paper, to overcome this computational limitation in the conventional approaches, a two-stage approach was proposed using the multiplicative dimensional reduction method (M-DRM) [24] and the standard gPC method [8]. In the proposed approach, the M-DRM was first used to detect important inputs by approximating a complicated function of random variables as a derivation of the univariate functions. The standard gPC method was then applied for UQ by considering the important inputs only detected in the SA step. The proposed two-stage approach was then illustrated on complex chemical processes, such as the propylene glycol production process and a lean

dry gas processing process. A rigorous simulation was performed by the interface between MATLAB™ and Aspen HYSYS™. The main target was to reduce the computational efforts (simulation time) for UQ significantly over conventional approaches, such as MC/QMC/standard gPC methods, when dealing with moderate/large number of uncertainties. The aim was to explain the efficient and practical framework and the typical steps involved SA and UQ.

2. Variance based-sensitivity analysis

Sensitivity analysis (SA) examines how the variability of the model output is affected by the uncertainty of various inputs. The methods for SA can be divided into two categories: local SA and global SA. The local approach uses one-factor-at-a-time experiments with gradients to study the local variability of the model, whereas global SA deals with the global variations in the output due to the given entire range of uncertainties on the inputs. This study focused on the global SA using Sobol's indices, which are also known as a variance based method, to determine the variables most responsible for the uncertainty in the model output.

Probability is a natural framework for modeling an uncertain input by assuming that the input is a *N*-dimensional random vector of mutually intendent components, $\boldsymbol{\xi} = (\xi_1, \xi_2, ..., \xi_N)$, with probability density functions of $\rho_i(\xi_i) : \Gamma_i \to R^+$;

A steady-state process, which is described conceptually in Fig. 1, was summarized in the form of nonlinear equations:

$$y = h(\boldsymbol{\xi}) \tag{1}$$

where $\mathbf{y}$ denotes a process output (quantity of interest).

The output mean and variance are defined as

$$\mu_y = \int_{\Gamma_1} \dots \int_{\Gamma_N} y(\xi_1, ..., \xi_N) \prod_{i=1}^{N} \rho_i(\xi_i) d\xi_i \quad \text{and} \tag{2}$$

$$D_y = \int_{\Gamma_1} \dots \int_{\Gamma_N} \left[(y(\xi_1, ..., \xi_N)) - \mu_y \right]^2 \prod_{i=1}^{N} \rho_i(\xi_i) d\xi_i \tag{3}$$

Global sensitivity analysis is based on a decomposition of the computational model in Eq. (1) as follows [25,26]:

$$y(\boldsymbol{\xi}) = y_0 + \sum_{i=1}^{N} y_i(\xi_i) + \sum_{i=1}^{N-1}\sum_{j>i}^{N} y_{ij}(\xi_i, \xi_j) + \quad + y_{i_1 \quad i_N}(\xi_i, \quad , \xi_N), \tag{4}$$

where $y_0 = \mu_y$.

The terms in Eq. (4) can be obtained recursively as follows:

$$\begin{aligned} y_i(\xi_i) &= \mathbf{E}\left[\mathrm{y}(\boldsymbol{\xi})\middle|\xi_i\right] - \mu_y \\ y_{ij}(\xi_i, \xi_j) &= \mathbf{E}\left[\mathrm{y}(\boldsymbol{\xi})\middle|\xi_i, \xi_j\right] - y_i - y_j - \mu_y, \end{aligned} \tag{5}$$

where $\mathbf{E}\left[y(\boldsymbol{\xi})\middle|[]\right]$ is the conditional expectation of $y(\boldsymbol{\xi})$ when the corresponding inputs are set. The components defined in (5) can be proven to be orthogonal with each other.

By taking advantage of the independent of the summands in Eq. (4), the output variance can be decomposed as:

$$D_y = \sum_{i=1}^{N} D_i + \sum_{i=1}^{N-1} \sum_{j>i}^{N} D_{ij} + \ldots + D_{1,2, \quad ,N} \tag{6}$$

where

$$\begin{aligned} D_i &= \mathrm{var}\left(\mathbf{E}\left[y(\boldsymbol{\xi})\middle|\xi_i\right]\right) \\ D_{ij} &= \mathrm{var}\left(\mathbf{E}\left[y(\boldsymbol{\xi})\middle|\xi_i, \xi_j\right]\right) - D_i - D_j \\ D_{1,2, \quad ,N} &= D_y - \sum_{i=1}^{N} D_i - \quad - \sum_{1 \le i_1 < \quad < i_{n-1} \le N} D_{i_1 \ldots i_{N-1}} \end{aligned} \tag{7}$$

In Eq. (7), the outer variance is taken over the corresponding inputs.

The first order Sobol's sensitivity index (function) quantifies the amount of the output variance that can be apportioned to the sole input variable, ξ_i:

$$S_i = \frac{D_i}{D_y}. \tag{8}$$

Similarly, higher order sensitivity functions describe what part of the total variance is due to the joint effect of inputs, $\{\xi_{i_1},...,\xi_{i_s}\}$, as

$$S_{i_1,...,i_k} = \frac{D_{i_1,...,i_k}}{D_y}. \tag{9}$$

The Sobol's total effect functions quantify the total impact of the factor, ξ_i, including all of its interactions with the other inputs:

$$T_i = \frac{\mathbf{E}(\mathrm{var}(y(\boldsymbol{\xi}) \mid \boldsymbol{\xi}_{\sim i}))}{D_y}. \tag{10}$$

In other words, if T_i is close to zero, the i[th] input, ξ_i, can be neglected.

The Monte Carlo method can be used to estimate the Sobol indices as follows [4].

- Generate a $\mathrm{A}_{Q\times N}$ matrix (Q is the sample size) from a given density function of inputs.

$$\mathrm{A}_{Q\times N} = \begin{bmatrix} \xi_1^{(1)} & \cdots & \xi_i^{(1)} & \cdots & \xi_N^{(1)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \xi_1^{(Q)} & \cdots & \xi_i^{(Q)} & \cdots & \xi_N^{(Q)} \end{bmatrix} \tag{11}$$

- Generate a $\mathrm{B}_{Q\times N}$ matrix (independent from A) from a given density function of inputs.

$$\mathrm{B}_{Q\times N} = \begin{bmatrix} \xi_1^{(Q+1)} & \cdots & \xi_i^{(Q+1)} & \cdots & \xi_N^{(Q+1)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \xi_1^{(2Q)} & \cdots & \xi_i^{(2Q)} & \cdots & \xi_N^{(2Q)} \end{bmatrix} \tag{12}$$

- Form matrices C_i from all columns of B except the i^{th} column, which is taken from A.
- Obtain vectors of the model output, $y_{\mathrm{A}} = M(\mathrm{A})$, $y_{\mathrm{B}} = M(\mathrm{B})$, $y_{\mathrm{C}_i} = M(\mathrm{C}_i)$, by computing the output of model (1) for all input values in (sample matrices) $\mathrm{A}, \mathrm{B}, \mathrm{C}_i$.
- First order indices are estimated as follows:

$$S_i = \frac{(1/Q)\sum_{j=1}^{Q} y_{\mathrm{A}}^{(j)} y_{\mathrm{C}_i}^{(j)} - M_0^{\,2}}{(1/Q)\sum_{j=1}^{Q} (y_{\mathrm{A}}^{(j)})^2 - M_0^{\,2}}, \tag{13}$$

where $M_0^2 = \left((1/Q) \sum_{i=1}^{Q} y_{\mathrm{A}}^{(j)} \right)^2$ is the empirical mean of the model output.

The total order indices is estimated as follows:

$$T_i = 1 - \frac{(1/Q) \sum_{j=1}^{Q} y_{\mathrm{B}}^{(j)} y_{\mathrm{C_i}}^{(j)} - M_0{}^2}{(1/Q) \sum_{j=1}^{Q} (y_{\mathrm{A}}^{(j)})^2 - M_0{}^2}. \tag{14}$$

For a model with N inputs, the total computational cost is $Q(N+2)$.

3. Sensitivity analysis using multiplicative dimensional reduction method

This section briefly describes the multiplicative dimensional reduction method (M-DRM) for the global sensitivity analysis from Zhang and Pandey [24]. The M-DRM method will be used to simplify the model for UQ with the polynomial chaos method.

3.1 Brief about M-DRM

Consider an overall response function, $y = h(\boldsymbol{\xi})$. Using the logarithmic transformation, one can obtain

$$\varphi(\boldsymbol{\xi}) = \log[abs(y)] = \log\{abs[h(\boldsymbol{\xi})]\}. \tag{15}$$

Following the univariate conventional dimensional reduction method (C-DRM) in the literature[27], an approximation of $\varphi(\boldsymbol{\xi})$ can be written as

$$\varphi(\boldsymbol{\xi}) \approx \sum_{i=1}^{N} \varphi(\xi_i, c_{-i}) - (N-1)\varphi_0, \tag{16}$$

where the functions are referred to those in the original space as follows:

$$\begin{cases} \varphi_0 = \log\{abs[h(c_1, c_2, ..., c_N)]\} \\ \varphi(\xi_i, c_{-i}) = \log\{abs[h(c_1, ..., c_{i-1}, \xi_i, c_{i+1}, c_N)]\} \end{cases}, \tag{17}$$

where $\mathbf{c} = (c_1, c_2, ..., c_N)$ is a cut point.

To invert the transformation, the original function can be written as

$$\begin{aligned}\exp[\varphi(\boldsymbol{\xi})] &\approx \exp\left[\sum_{i=1}^{N}\varphi(c_1,...,c_{i-1},\xi_i,c_{i+1},c_N)-(N-1)\varphi_0\right] \\ &= \exp[(1-N)\varphi_0]\times\exp\left[\sum_{i=1}^{N}\varphi(c_1,...,c_{i-1},\xi_i,c_{i+1},c_N)\right]\end{aligned} \quad (18)$$

Substituting for the expressions from Eq. (17) into Eq. (18) leads to a multiplicative approximate of the response function:

$$h(\boldsymbol{\xi}) \approx [h(c)^{1-N}].\prod_{i=1}^{N} h(c_1,...,c_{i-1},\xi_i,c_{i+1},c_N) = h_0^{1-N}\prod_{i=1}^{N} h_i(\xi_i,c_{-i}) \quad (19)$$

This approximate model of the original input-output relation is known as the univariate M-DRM.

3.2 Variance based sensitivity analysis by the M-DRM

Denote the mean and mean square of the k the dimensional function as

$$\begin{cases}\rho_k = \mathbf{E}\left[h_k(\xi_k)\right] \\ \theta_k = \mathbf{E}\left[h_k(\xi_k)^2\right]\end{cases} \quad (20)$$

The mean and mean square of the output can be approximated as

$$\mu_y \approx h_o^{1-N}\prod_{k=1}^{N}\rho_k \;\; ;\mathbf{E}\left[y^2\right] \approx h_0^{2-2N}\prod_{k=1}^{N}\theta_k \quad (21)$$

Under the M-DRM approximation, an i^{th} conditional expectation function, D_i, can be evaluated as

$$\mathbf{E}\left[y(\boldsymbol{\xi})|\xi_i\right] \approx \mathbf{E}\left[h_0^{1-N}.\prod_{k=1}^{N}h_k(\xi_k)\right] = h_0^{1-N}.h_i(\xi_i).\prod_{k=1,k\neq i}^{N}\rho_k \quad (22)$$

The conditional second moment can be derived as

$$\mathbf{E}\left[y^2(\boldsymbol{\xi})|\xi_i\right] \approx \int_{\xi_i}\left(h_0^{1-N}.h_i(\xi_i).\prod_{k=1,k\neq i}^{N}\rho_k\right)^2 f_i(\xi_i)dx_i = h_0^{2-2N}.\theta_i.\prod_{k=1;k\neq i}^{N}\rho_k^2 = \mu_Y^2.\theta_i/\rho_i^2 \quad (23)$$

Therefore, the primary variance can be approximated as

$$D_i = \mathrm{var}\left(\mathbf{E}\left[y(\boldsymbol{\xi})|\xi_i\right]\right) = \mu_y{}^2(\theta_i/\rho_i^2-1) \quad (24)$$

The first order index can be approximated as

$$S_i = \frac{D_i}{D_Y} \approx \frac{\theta_i / \rho_i^2 - 1}{\left(\prod_{k=1}^{N} \theta_k / \rho_k^2 \right) - 1} \tag{25}$$

To evaluate the total sensitivity index, T_i, it is essential to calculate the following conditional variance:

$$\mathrm{var}(y(\boldsymbol{\xi}) \mid \boldsymbol{\xi}_{\sim i}) = \mathbf{E}((y(\boldsymbol{\xi}) \mid \boldsymbol{\xi}_{\sim i})^2) - \mathbf{E}((y(\boldsymbol{\xi}) \mid \boldsymbol{\xi}_{\sim i}))^2 \tag{26}$$

Owing to the M-DRM, the expectations in the right hand side of Eq. (26) can be approximated as

$$\mathbf{E}((y(\boldsymbol{\xi}) \mid \boldsymbol{\xi}_{\sim i})^2) \approx h_0^{2-2N} . \left(\prod_{k=1, k \neq i}^{N} [h_k(\xi_k)]^2 \right) . \theta_i \tag{27}$$

$$\mathbf{E}((y(\boldsymbol{\xi}) \mid \boldsymbol{\xi}_{\sim i}))^2 \approx h_0^{2-2N} . \left(\prod_{k=1, k \neq i}^{N} [h_k(\xi_k)]^2 \right) . \rho_i^2 \tag{28}$$

The conditional variance is obtained as

$$\mathrm{var}(y(\boldsymbol{\xi}) \mid \boldsymbol{\xi}_{\sim i}) \approx h_0^{2-2N} . \left(\prod_{k=1, k \neq i}^{N} \theta_k \right) . (\theta_i - \rho_i^2) \tag{29}$$

The expectation of the conditional variance is obtained as

$$\mathbf{E}(\mathrm{var}(y(\boldsymbol{\xi}) \mid \boldsymbol{\xi}_{\sim i})) \approx h_0^{2-2N} . \left(\prod_{k=1, k \neq i}^{N} \theta_k \right) . (\theta_i - \rho_i^2) \tag{30}$$

Finally, the total sensitivity index of ξ_i can be approximated as

$$T_i \approx \frac{1 - \rho_i^2 / \theta_i}{1 - \left(\prod_{k=1}^{N} \rho_k^2 / \theta_k \right)} \tag{31}$$

Furthermore, the Gaussian quadrature with q nodes is moderately efficient for calculating the one-dimensional integrals in an integration of the univariate functions as follows:

$$\begin{cases} \rho_k \approx \sum_{l=1}^{q} w_{kl} h(c_1, ..., c_{k-1}, \xi_k^{(l)}, c_{k+1}, ..., c_N) \\ \theta_k \approx \sum_{l=1}^{q} w_{kl} [h(c_1, ..., c_{k-1}, \xi_k^{(l)}, c_{k+1}, ..., c_N)]^2 \end{cases} \tag{32}$$

Regarding the standard of the M-DRM approximation, only Nq total number of functional evaluations are required for complete sensitivity analysis. Note that only a small number of nodes q (5-10) is normally used in Gaussian quadrature. On the other hand, the MC method indices requires $(N+2)Q$ to estimate the sensitivity, where Q is a large number of samples (e.g., 10000).

4. Uncertainty quantification with polynomial chaos

Assume that $n<N$ important inputs are detected from SA in the previous step. The following polynomial chaos based method[8] can be used for UQ. The unimportant inputs are fixed to their nominal value. Therefore, the response function is expanded into a series of n variate P^{th} order polynomials:

$$\begin{aligned} y(\boldsymbol{\xi}) &= \sum_{i=1}^{M} f_i \Phi_i(\boldsymbol{\xi}); \\ M+1 &= \binom{n+P}{n} = \frac{(n+P)!}{n!P!} \end{aligned} \tag{33}$$

where f_m is the coefficient of gPC expansion that satisfies

$$f_i = \mathbf{E}[\Phi_i f(y)] = \int_{\Gamma} f(y)\Phi_i(\boldsymbol{\xi})\rho(\boldsymbol{\xi})d\boldsymbol{\xi}. \tag{34}$$

In Eq. (34), these coefficients of the gPC expansion can be calculated numerically through a discrete projection according to the procedure reported by Xiu [28]:

- Choose a n-dimensional integration rule with $q_1 \times \ldots \times q_n$ cubature nodes/weights

$$^{q_1 \times \ldots \times q_n}[g] = \sum_{j_1=1}^{q_1} \ldots \sum_{j_N=1}^{q_n} g(\xi_1^{(j_1)}, \ldots, \xi_n^{(j_n)})(w_1^{(j_1)} \ldots w_1^{(j_n)}) \cong \int_{\Gamma} g(\boldsymbol{\xi})\rho(\boldsymbol{\xi})d\boldsymbol{\xi} \tag{35}$$

where $^{q_1 \times \ldots \times q_n}[\cdot]$ refers to the cubature numerical integration.

- Using the numerical integration rule in Eq. (35), the gPC coefficients can be approximated as:

$$f_j = {}^{q_1 \times \ldots \times q_n}[f(y,\boldsymbol{\xi})\Phi(\boldsymbol{\xi})\boldsymbol{\rho}] = \sum_{m=1}^{q_1 \times \ldots \times q_n} f(\boldsymbol{\xi}^{(m)})\Phi_j(\boldsymbol{\xi}^{(m)})w^{(m)} \quad \text{for} \quad j = 1, \ldots, M \tag{36},$$

where f_j is approximated numerically by f_j and $f(\xi)\Phi_j(\xi)$ plays a role of $g(\xi)$ in Eq. (35). The number of nodes (simulation) in the cubature rule rises exponentially. On the other hand, because only the important inputs (3-5 inputs) are considered, the number of simulations is still acceptable. An n-variate P^{th} order gPC approximation of the response function can be constructed in the form,

$$f_N^P = \sum_{j=1}^{M} f_j \Phi_j(\xi). \tag{37}$$

The truncated gPC expansion (37) contains all information about the statistical properties of the random output. Owing to the orthonormality of the gPC polynomials, the mean and variance of output may be calculated directly from the gPC coefficients.

The mean value is the first expansion coefficient:

$$\mu_y = \int_\Gamma f_N^P \rho(\xi) d\xi = \int_\Gamma \left[\sum_{j=1}^{M} f_j(\xi)\Phi_j(\xi) \right] \rho(\xi) d\xi = f_1. \tag{38}$$

The variance of the output $y(\xi)$ can calculated as the sum of the square of the remaining coefficients:

$$D_y = \sigma^2{}_y = \mathbf{E}[(y-\mu_y)^2] = \int_\Gamma (\sum_{j=1}^{M} f_j(\xi)\Phi_j(\xi) - f_1)(\sum_{j=1}^{M} f_j(\xi)\Phi_j(\xi) - f_1)\rho(\xi)d\xi = \sum_{j=2}^{M} f_j{}^2. \tag{39}$$

The output density function can be obtained by sampling the cheap-to-evaluate truncated gPC model in Eq. (37).

5. Case study

In this study, the UQ for two complex chemical processes was estimated based on the two-stage gPC-based approach. The emphasis is to explain the efficient, practical framework, and the principal steps involved SA and UQ, while further analysis can lead to a better understanding of those process models.

5.1. Example 1: propylene glycol production process with six uniform uncertainties

Propylene glycol (PG) has been used widely in industries, such as a food grade coolant in food industry[29], solvent in cosmetics[30], and de-icing fluid in aviation[31]. Referring to the conceptual model from HYSYS™,

Fig. 2 presents a flow diagram of a PG production process. In this process, propylene oxide (PO) is combined with water to produce PG in a continuously-stirred-tank reactor (CSTR). Owing to the exothermic reaction, a coolant stream circulates within the reactor jacket to remove any extra heat. The outlet stream is then introduced to a distillation column, in which PG is essentially recovered from the bottom stream with a purity of 99.5 wt. %. The distillation column operating at atmospheric pressure has 10 stages with a full reflux condenser and reboiler.

Adapted from a practical point of view, the variations in temperature, feed flowrate, pressure, and other properties should be considered. In this study, the flowrates of PO and water, the temperature and pressure of the mixed stream, the temperature of the reactor effluent, and the reflux ratio of the column were assumed to be independently uncertain and distributed uniformly in intervals of. Other parameters, such as the reactor vessel volume, column pressure, and number of theoretical stages were assumed to be deterministic. According to the M-DRM, a set of 60 Gaussian quadrature nodes ,which was generated using the MATLAB^{TM} codes from the orthogonal polynomial toolbox [32], was passed to HYSYS^{TM}, where the PG process in Fig. 2 was modeled rigorously. The reboiler duty as outputs from HYSYS^{TM} were collected and used for SA in Eq. (31). The Sobol's sensitivity indices, which can be used to identify the inputs of the model and have the largest influence on the prediction, were derived for the UQ and SA. Table 2 lists the sensitivity indices obtained from the M-DRM. As a result, the two random variables, i.e., the flowrate of water and the reflux ratio, were detected as being important while the heat duty was most sensitive to the flowrate of the feed water. Other random variables, including the PO feed flow rate, the temperature and pressure of the mixed stream, and the outlet temperature of the reactor effluent become non-influential factors that can be omitted in UQ. In particular, owing to the effective detection of the non-influential input of the M-DRM, one can simplify the model, and a 10^{th} order with two variates gPC, which requires only 10x10 simulations, was constructed for the UQ of the process. Because precise estimates of the process output cannot be accessible, the results of the proposed method were compared with those of the QMC method with a sufficiently large number of simulations. For an accurate estimation of the probability, the number of samples for the QMC

method was selected from Chernoff bound in Tempo et al.[33]. Fig. 3 compares the density functions obtained using the proposed and QMC methods with two influential random inputs (water flow rate and reflux ratio) and that by the QMC method with all six random inputs (using the 10000 simulations from Halton sequence). Table 1 lists the statistical properties of the reboiler duty of the distillation column (Q) achieved from the proposed and QMC methods. Table 1 also lists the computational cost required for the two methods. The computational time for the proposed method includes the computational time for solving the number of functional evaluations and performing simulations from SA and UQ. The QMC/MC methods require a large number of simulations to estimate the expected values, variances, and density accurately; hence, they are computationally expensive.

***5.2. Example* 2**: a lean dry gas processing plant with six uniform uncertainties

Fig. 4 shows a process schematic diagram of a lean dry gas production plant [34]. Two natural gas streams containing N_2, CO_2, and C_1 - n-C_4 hydrocarbons were mixed together and then processed in a refrigerated cycle to eliminate the heavier components. Ensuring the safe transport and processing of natural gas is crucial because hydrocarbon liquid dropout in gas transmission lines normally leads to a number of problems, including increased pressure drops, reduced pipeline capacity, and equipment problems, such as compressor damage [1,35,36]. In this case study, a mixed feed stream from two different sources is introduced into an inlet separator to remove the free hydrocarbon liquids. The overhead outlet gas from the separator is then entered into a low-temperature separator through two heat exchangers (gas/gas heat exchanger and chiller), in which the required cooling is accomplished. In the low-temperature separator, heavy hydrocarbons are removed so that the eventual sales gas meets the required properties, such as the heating values and dew point. The lean dry gas from the separator is fed to the gas/gas heat exchanger and then sent to the consumers through a pipeline network. The liquid streams coming from both separators are mixed and introduced to the depropanizer column to recover the propane and remaining heavy components.

This case study focuses mostly on determining the effects of the decision factors according to the operability limits for each of the uncertain variables. An interface under the code developed in MATLAB®

was used to make a random generation of the process inputs and quantify the uncertainty performance of the process inputs. First, Sobol's indices via SA were derived to detect the influential factors on the output variations. In this case study, the sale gas heating value was controlled under uncertainties while the flow rates (*F1* and *F2*), temperature (*Tn*), pressure (*Pn*) of two feed natural gas streams, outlet temperature of the cold gas (*Tc*), and reflux ratio of the column (*R*) were assumed to be uncertain independently by the uniform distribution in the range, such as $F1 \in [1.90; 2.32$ kg/s], $F2 \in [1.25; 1.52$ kg/s], $Tn \in [14.0; 17.1$ °C], $Pn \in [37.2; 45.5$ bars], $Tc \in [-16.8; -13.8$ °C], and $R \in [0.9; 1.1]$. The QMC method was also examined to determine the impact of these uncertainties on the process. Table 1 lists the statistical properties obtained from the proposed (M-DRM+gPC) and QMC methods for the heating value of the lean gas. In addition, Table 2 provides the Sobol's indices obtained by the M-DRM approximation. This indicates that the pressure of the feed natural gas and the outlet temperature of the cold gas are two influential factors for uncertainty propagation, while the other parameters can be fixed as their nominal values. At the second stage, the standard gPC approach can be used for UQ with only two random inputs. Fig. 5 compares the performance for the density functions of the lean gas heating value with two decision random inputs (the pressure of the feed natural gas and the output temperature of the cold gas) using two methods: the red line denotes the standard gPC method (with 100 samples) and the blue one is the QMC method (with 10000 samples). The results from the gPC and QMC methods using the two random inputs showed good agreement with that of the QMC method using all six random parameters with 10000 simulations (a green line). The proposed M-DRM approach detected the influential parameter inputs correctly through the derived SA indices.

Remark Owing to the exponential increase in simulation efforts, the standard gPC method was not considered in Examples 1 and 2 for cases with six random inputs, and only the M-DRM and QMC methods for UQ were considered. In addition, the QMC method was not used for the SA of the two chemical processes

because of the large computational effort required for SA using the QMC method (This will be approximately eight times higher than the effort for UQ).

5. Conclusions

The purposes of UQ and SA in process design and modeling are important for testing model robustness against uncertainties, obtaining a better understanding of the effects of inputs on the output, determining the key inputs that mainly influence the uncertainty of model output, and achieving a model simplification. In this study, to overcome the computational limitation in the UQ and SA of complex chemical processes, a two stage gPC approach was presented based on M-DRM and gPC for UQ and SA. In the proposed method, the SA indices were first estimated by the M-DRM in a computationally efficient manner and the gPC method was then applied to the UQ based on the simplified model from the M-DRM. HYSYS^{TM} was used to obtain a rigorous simulation result. The M-DRM approximated the complex uncertain problem successfully in the form of the product of univariate functions using a small number of sampling points. Through the M-DRM approximation, the Sobol's sensitivity indices for detecting non-influential inputs could be obtained with little computational burden. The proposed two-stage approach is superior to the popular QMC/gPC approaches, primarily in terms of the computational efforts when a large number of random inputs are considered. The proposed approach was applied to the UQ and SA of the propylene glycol production process and the lean dry gas processing plant. The results showed precise agreement with those of the conventional QMC method, of which the computational cost is in the order of 10^4 model evaluations, and not compatible with the UQ and SA of most chemical processes with a moderate/large number of uncertainties.

ACKNOWLEDGMENTS

This study was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2015R1D1A3A01015621) and by the Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2014R1A6A1031189). Pham L.T. Duong and Jorge Goncalves were supported by FNR CORE project ref 8231540.

Reference

1. Abubakar U, Sriramula S, Renton NC. Reliability of complex chemical engineering processes. *Comput. Chem. Eng.* 2015; 74:1-14.

2. Lucay F, Cisternas LA, Gálvez ED. Global sensitivity analysis for identifying critical process design decisions. *Chem. Eng. Res. Des.* 2015; 103:74-83.

3. O'Hagan A. Bayesian analysis of computer code outputs: A tutorial. *Reliab. Eng. Syst. Saf.* 2006; 91:1290-1300.

4. Binder K. Monte Carlo Methods: a powerful tool of statistical physics. In: Niederreiter H, Hellekalek P, Larcher G, Zinterhof P, eds. *Monte Carlo and Quasi-Monte Carlo Methods 1996: Proceedings of a conference at the University of Salzburg, Austria, July 9–12, 1996*. New York, Springer; 1998:19-39.

5. Coulibaly I, Lécot C. Monte Carlo and quasi-Monte Carlo algorithms for a linear integro-differential equation. In: Niederreiter H, Hellekalek P, Larcher G, Zinterhof P, eds. *Monte Carlo and Quasi-Monte Carlo Methods 1996: Proceedings of a conference at the University of Salzburg, Austria, July 9–12, 1996*. New York, Springer; 1998:176-188.

6. Liu JS. Sequential Monte Carlo in Action. *Monte Carlo Strategies in Scientific Computing*. New York, Springer; 2004:79-104.

7. Kroese DP, Taimre T, Botev ZI. Markov Chain Monte Carlo. *Handbook of Monte Carlo Methods*: John Wiley & Sons, Inc.; 2011:225-280.

8. Duong PLT, Ali W, Kwok E, Lee M. Uncertainty quantification and global sensitivity analysis of complex chemical process using a generalized polynomial chaos approach. *Comput. Chem. Eng.* 2016; 90:23-30.

9. Baroni G, Tarantola S. A General Probabilistic Framework for uncertainty and global sensitivity analysis of deterministic models: A hydrological case study. *Environ. Model. Softw.* 2014; 51:26-34.

10. Saltelli A, Tarantola S, Campolongo F, Ratto M. Global Sensitivity Analysis for Importance Assessment. *Sensitivity Analysis in Practice*: John Wiley & Sons, Ltd; 2004:31-61.

11. Saltelli A, Chan K, Scott EM. *Sensitivity analysis*: Wiley, New York; 2000.

12. Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S. *Global sensitivity analysis: the primer*: John Wiley & Sons; 2008.

13. Celse B, Costa V, Wahl F, Verstraete JJ. Dealing with uncertainties: Sensitivity analysis of vacuum gas oil hydrotreatment. *Chem. Eng. J.* 2015; 278:469-478.

14. Papalambros PY, Wilde. DJ. *Principles of Optimal Design*: Cambridge University Press; 2000.

15. Chung PS, Jhon MS, Biegler LT. Chapter 2 - The Holistic Strategy in Multi-Scale Modeling. In: Guy BM, ed. *Advances in Chemical Engineering.* Vol. 40: Academic Press; 2011:59-118.

16. Wiener N. The homogeneous chaos. *Amer. J. Math.* 1938; 60(4):897-936.

17. Nagy ZK, Braatz RD. Distributional uncertainty analysis using power series and polynomial chaos expansions. *J. Process Control.* 2007; 17(3):229-240.

18. Shen DE, Braatz RD. Polynomial chaos-based robust design of systems with probabilistic uncertainties. *AlChE J.* 2016; 62(9):3310-3318.

19. Duong PLT, Lee M. Robust PID controller design for processes with stochastic parametric uncertainties. *J. Process Control.* 2012; 22(9):1559-1566.

20. Duong PLT, Lee M. Probabilistic analysis and control of systems with uncertain parameters over non-hypercube domain. *J. Process Control.* 2014; 24(4):358-367.

21. Du YC, Duever TA, Budman H. Fault detection and diagnosis with parametric uncertainty using generalized polynomial chaos. *Comput. Chem. Eng.* 2015; 76:63-75.

22. Xiu D, Karniadakis GE. The Wiener--Askey Polynomial Chaos for Stochastic Differential Equations. *SIAM J. Sci. Comput.* 2002; 24(2):619-644.

23. Ghanem RG, Spanos PD. *Stochastic finite elements: a spectral approach*: Courier Corporation; 2003.

24. Zhang X, Pandey MD. An effective approximation for variance-based global sensitivity analysis. *Reliab. Eng. Syst. Saf.* 2014; 121:164-174.

25. Saltelli A, Tarantola S, Campolongo F, Ratto M. Sensitivity Analysis in Diagnostic Modelling: Monte Carlo Filtering and Regionalised Sensitivity Analysis, Bayesian Uncertainty Estimation and Global Sensitivity Analysis. *Sensitivity Analysis in Practice*: John Wiley & Sons, Ltd; 2004:151-192.

26. Saltelli A, Tarantola S, Campolongo F, Ratto M. Methods Based on Decomposing the Variance of the Output. *Sensitivity Analysis in Practice*: John Wiley & Sons, Ltd; 2004:109-149.

27. Xu H, Rahman S. A generalized dimension-reduction method for multidimensional integration in stochastic mechanics. *Int. J. Numer. Meth. Eng.* 2004; 61(12):1992-2019.

28. Xiu D. *Numerical Methods for Stochastic Computations: A Spectral Method Approach*: Princeton University Press; 2010.

29. Smolinske SC. *Handbook of food, drug, and cosmetic excipients*. Boca Raton: CRC Press; 1999.

30. Sullivan CJ. Propanediols. *Ullmann's Encyclopedia of Industrial Chemistry*: Wiley-VCH Verlag GmbH & Co. KGaA; 2000.

31. Bausmith DS, Neufeld RD. Soil Biodegradation of Propylene Glycol Based Aircraft Deicing Fluids. *Water Environ. Res.* 1999; 71(4):459-464.

32. Gautschi W. *Orthogonal Polynomials: Computation and Approximation*: Clarendon Press; 2004.

33. Tempo R, Calafiore G, Dabbene F. *Randomized algorithms for analysis and control of uncertain systems: with applications*: Springer Science & Business Media; 2012.

34. AspenTech. Aspen Physical Property System. 2009; http://www.aspentech.com/products/engineering/aspen-properties/. Accessed 07.09.2016.

35. Mørch Ø, Nasrifar K, Bolland O, Solbraa E, Fredheim AO, Gjertsen LH. Measurement and modeling of hydrocarbon dew points for five synthetic natural gas mixtures. *Fluid Phase Equilib.* 2006; 239(2):138-145.

36. Skylogianni E, Novak N, Louli V, Pappa G, Boukouvalas C, Skouras S, Solbraa E, Voutsas E. Measurement and prediction of dew points of six natural gases. *Fluid Phase Equilib.* 2016; 424:8-15.

List of Figures

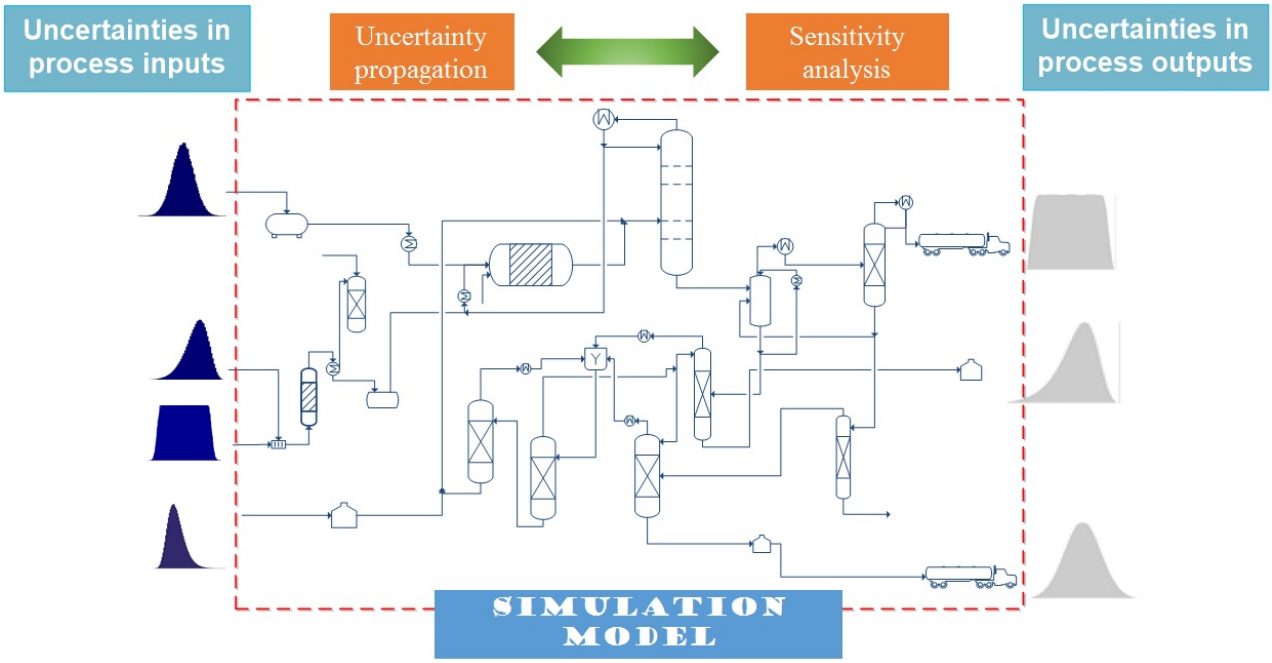

Fig. 1. Concept of uncertainty analysis in chemical processes.

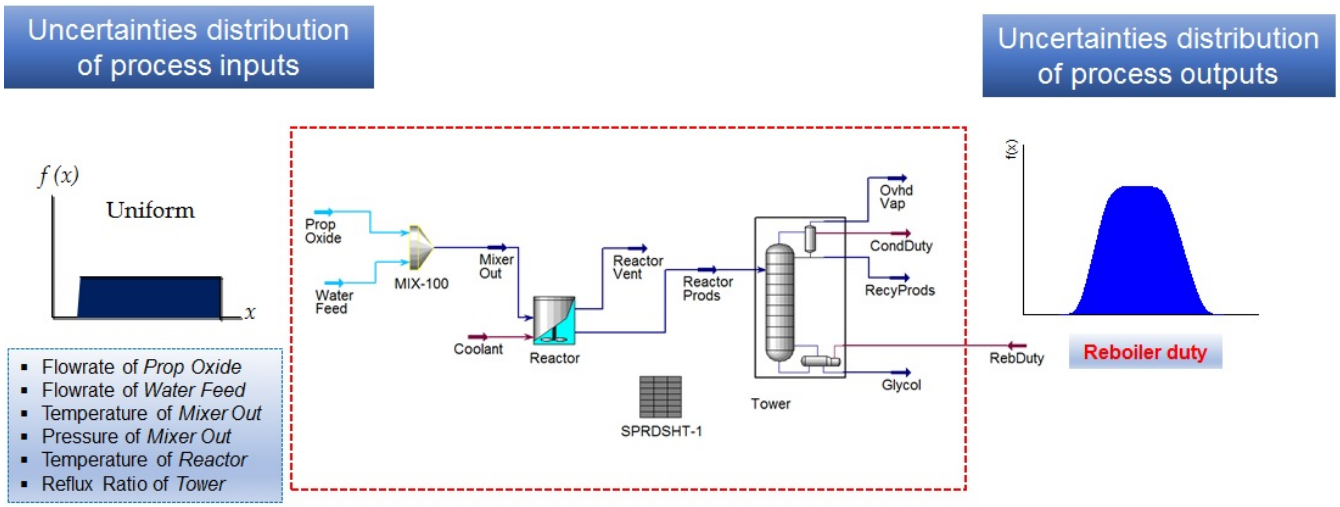

Fig. 2. Propylene glycol (PG) production process (Example 1).

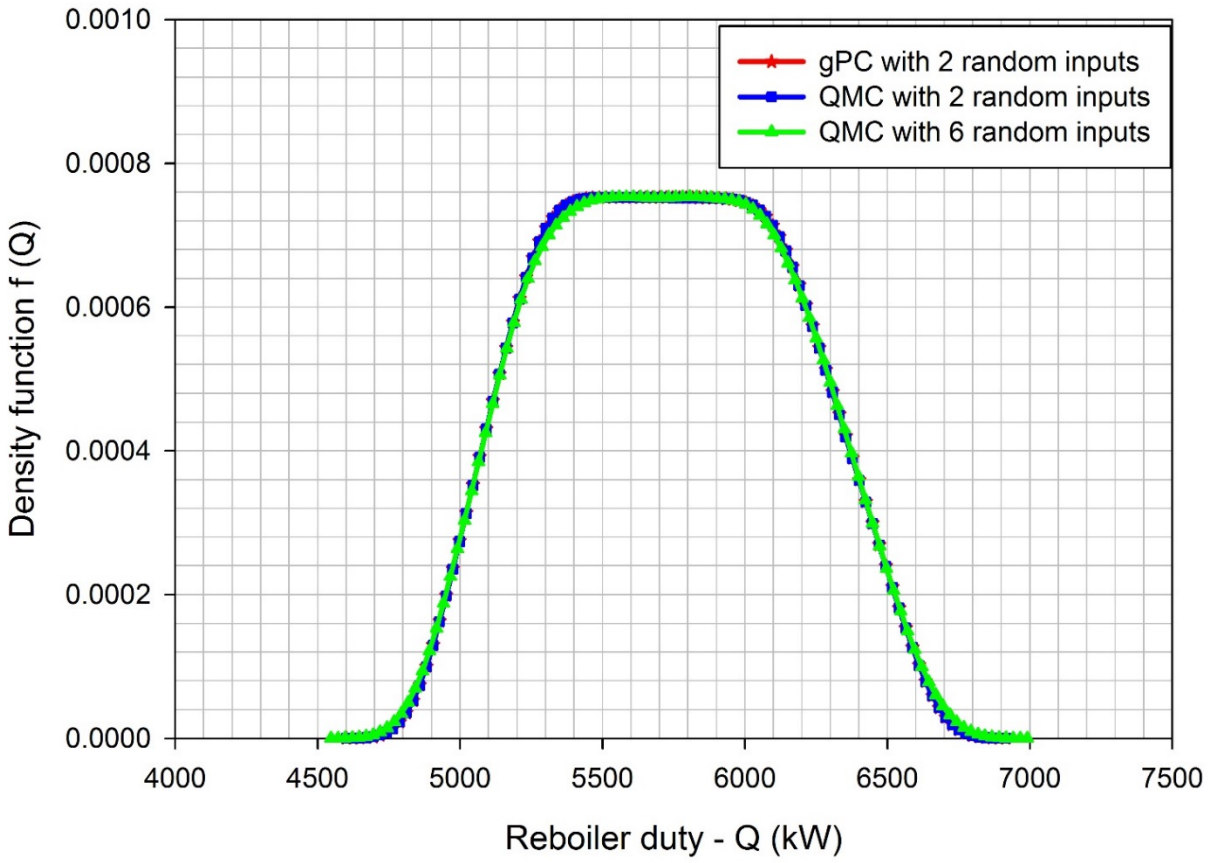

Fig. 3. Density distributions of the reboiler duty in the PG production process with six and two uniform random inputs (Example 1).

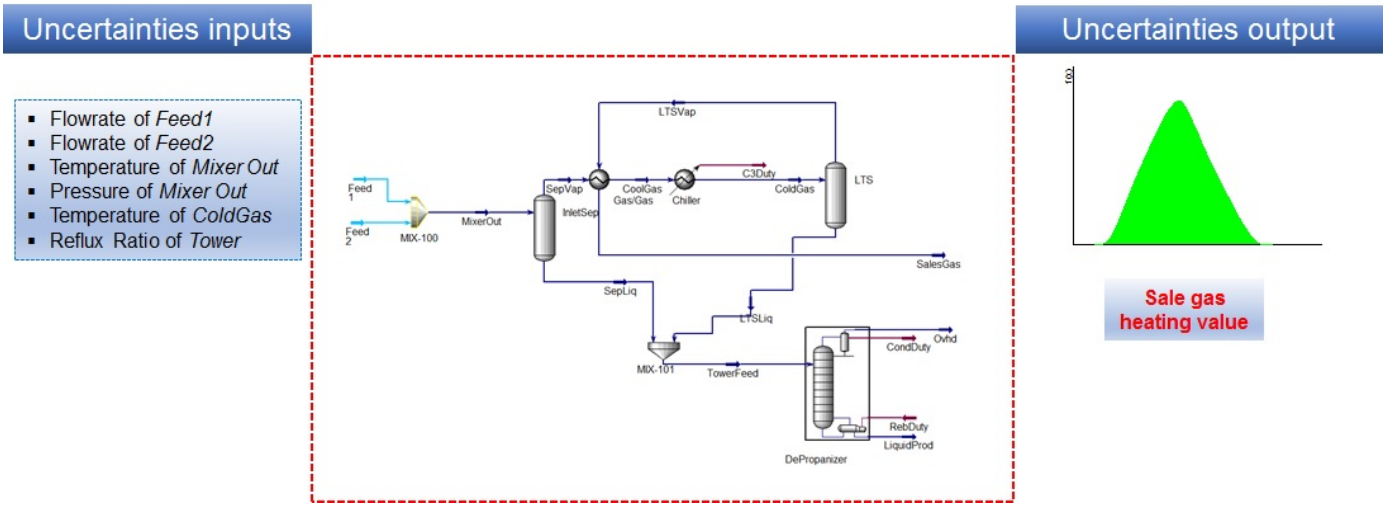

Fig. 4. Lean dry gas processing plant (Example 2).

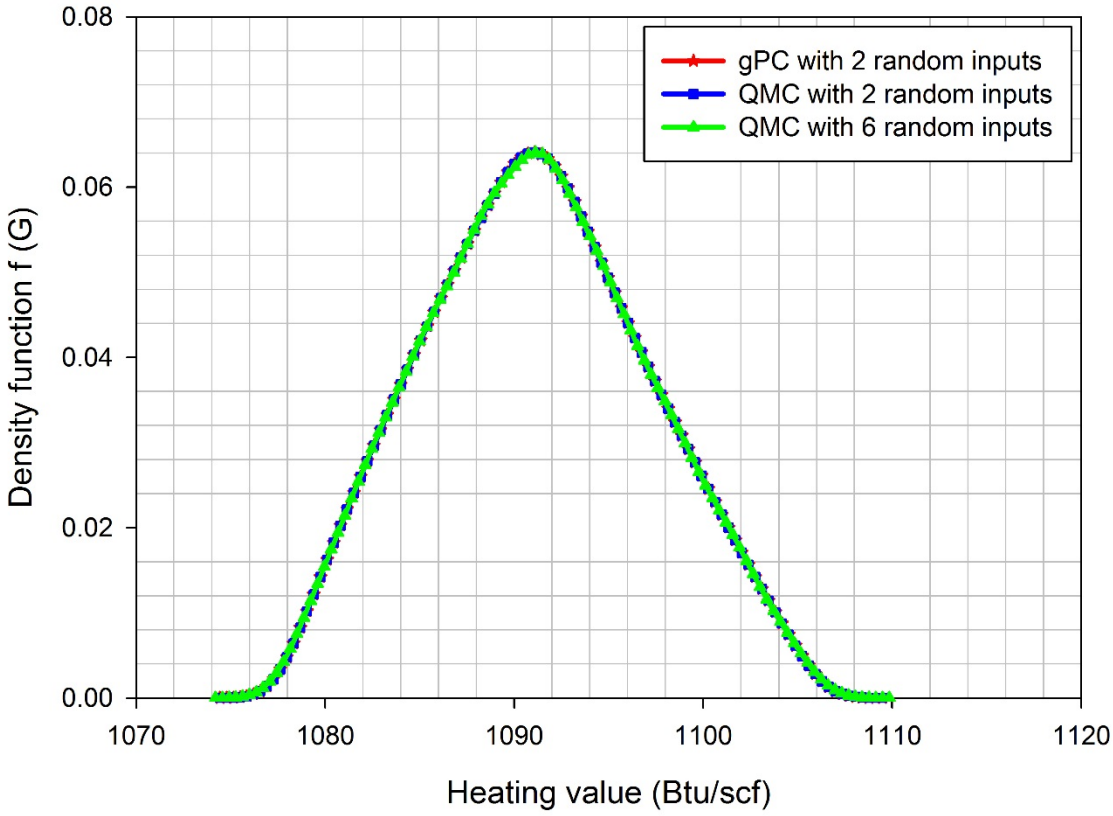

Fig. 5. Density distributions of the net heating value of sale gas in the lean dry gas production process with six and two uniform random inputs (Example 2).

List of Tables

Table 1. Simulation parameters and computational time profiles for obtaining the statistical characteristics using the MDRM-gPC/QMC methods for Examples 1 and 2 (with six random inputs).

Method	Example 1			Example 2		
	No. of simulations	**Runtime (sec.)**	**Mean μ(Q)**	**No. of simulations**	**Runtime (sec.)**	**Mean μ(BTU/SCF)**
QMC	10000	19738.6	5729.4	10000	19693.8	1091.1
Proposed (M-DRM+gPC)	160	230.5	5729.2	160	228.3	1090.9

Table 2. Sobol's sensitivity indices from the surrogated model by M-DRM for Examples 1 and 2.

	Sobol's Sensitivity Indices (S_i, T_i)							
Example 1	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**T1**	**T2**
	0.0104	0.8532	6.92e-08	1.52e-11	0.0097	0.1260	0.0104	0.8539
	T3	**T4**	**T5**	**T6**				
	9.96e-08	1.53e-11	0.0097	0.1266				
Example 2	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**T1**	**T2**
	4.58e-04	4.57e-04	0.0038	0.5526	0.4427	7.46e-12	4.58e-04	4.57e-04
	T3	**T4**	**T5**	**T6**				
	0.0038	0.5526	0.4427	7.46e-12				