Reliability Engineering and System Safety xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Efficient sensitivity analysis and interpretation of parameter correlations in chemical engineering

Xiangzhong Xie[a,b,c], René Schenkendorf[*,a,b], Ulrike Krewer[a,b]

[a] Institute of Energy and Process Systems Engineering, Braunschweig University of Technology, Franz-Liszt-Straße 35, Braunschweig 38106, Germany
[b] Center of Pharmaceutical Engineering (PVZ), Braunschweig University of Technology, Franz-Liszt-Straße 35a, Braunschweig 38106, Germany
[c] International Max Planck Research School (IMPRS) for Advanced Methods in Process and Systems Engineering, Sandtorstraße 1, Magdeburg 39106, Germany

ARTICLE INFO

Keywords:
Sensitivity analysis
Polynomial chaos expansion
Parameter correlation
Gaussian copula
Chemical processes

ABSTRACT

Parameter uncertainties affect model-based system reliability analysis and may lead to safety issues in model-based process design. Global sensitivity analysis (GSA) is a valuable tool to quantify the influence of parameter uncertainties in the variation of the model output. However, GSA has not been widely employed in the field of chemical engineering, especially for processes with correlated model parameters. Parameter correlations, in turn, are quite common when identifying model parameters with experimental data. Thus, we propose and critically compare (co)variance-based and moment-independent GSA techniques for analyzing chemical processes in the absence and presence of parameter correlations. Technically, polynomial chaos expansion is used to reduce the computational burden for GSA. The proposed methods are demonstrated for a continuous synthesis process. Here, the results show significant differences in the parameter sensitivity rankings when parameter correlations are considered or not while the moment-independent technique provides a universal and easy-to-interpret sensitivity measure.

1. Introduction

Engineers use optimization and control tools based on mathematical models to design the structure and improve the reliability and safety of chemical processes [2,14,28]. Mathematical models, which attempt to mimic the chemical processes, are actually functions of parameters. However, different types of uncertainties are present in the model parameters, and thus, the model outputs deviate from reality and vary within a certain range. The existence of uncertainties increases the difficulty in model-based design of chemical processes [55,63] and may lead to critical operation conditions. To achieve reliable model-based results, methods for quantifying and analyzing the uncertainties of chemical processes are required. The parameter uncertainties usually result from either the imperfection of experimental measurements (epistemic uncertainties) or the intrinsic randomness (aleatoric uncertainties). The uncertainties can be decreased but not totally eliminated by adding more information from experiments [3]. Many scholars have investigated the design of chemical processes under the condition of uncertain parameters [18,26,38,49]. These studies are based on propagating all parameter uncertainties through the chemical processes and quantifying the variation in the model response. The aim of the works is to reduce the variation in the model response and robustify the processes by using systematic design tools. However, as the computational demand for embedding uncertainty propagation and quantification (UQ) into process design tools increases exponentially with the number of uncertain parameters [3], their efficiency is not guaranteed for large-scale problems. One solution for this is to screen out the parameters that have significant impacts on the model responses and reduce the scale of the problems by neglecting the rest of the parameters [40]. The method used to analyze the influence of uncertain parameters on the model response is called sensitivity analysis (SA). In practice, even in the recent past, SA has been falsely implemented and misinterpreted in various disciplines, especially in the field of chemical engineering [39]. As done in this work, it is necessary to demonstrate the impact of critical factors, e.g., the existence of parameter correlations, on the reliability of the results of SA and the way how to deal with them appropriately.

SA plays an essential role in mathematical modeling and model-based design, which allows us to have a deep insight into the relevance of model parameters and model responses [6,10,44]. SA provides crucial information for model reduction [52] and control strategy design [20]. SA methods are categorized into two groups: local and global SA. Local SA quantifies deviation of the model response with respect to a step change in an individual parameter around its nominal value. The

* Corresponding author at: Institute of Energy and Process Systems Engineering, Braunschweig University of Technology, Franz-Liszt-Straße 35, Braunschweig 38106, Germany.
E-mail address: r.schenkendorf@tu-braunschweig.de (R. Schenkendorf).

https://doi.org/10.1016/j.ress.2018.06.010
Received 12 January 2018; Received in revised form 24 May 2018; Accepted 17 June 2018

associated methods for local SA are the finite difference-based method [6] and differential-based method [5]. The major advantage of local SA methods is that they are simple and straightforward to implement without cumbersome model evaluations. However, local SA methods lead to the inevitable loss of information regarding the sensitivities and interactions of the parameters because only the local and individual effects of parameters are considered [41]. In contrast, global SA (GSA) quantifies variations in model output on the entire domain of the parameter space and comprehensively analyzes the interactions among parameters. Different techniques for GSA are available, such as the derivative-based methods [9,48], non-parametric methods [12,17], variance-based methods [42,43], and moment-independent methods [4]. A detailed review of those methods is provided by Saltelli et al. [6,40]. Numerous works, which implement SA for chemical processes, are based on local SA, such as [1,20,52]. However, [16,64] compared the results from local SA and GSA for a chemical reactor design and concluded that GSA is preferable, especially for chemical processes with high nonlinearity. In this article, two highly promising methods for GSA are compared: (i) Sobol' sensitivity indices (SSI) representing the classical variance-based approach [46], and (ii) the moment-independent sensitivity analysis (MISA) analyzing the entire model variation [4].

The SSI and MISA are commonly used for sensitivity analysis of problems with independent model parameters [3,4,8]. However, problems with correlated model parameters, which arise from inherent parameter dependences or parameter identification procedures, are the standard in chemical processes and other industrial applications [16,27,36,53]. The parameter space is restricted by the correlations and is different from the one without correlations, which is illustrated in Fig. 1. This could lead to a different impact onto the analyzed process model. Thus, the sensitivity analysis and robust design under the hypothesis of independent parameters might lead to unreliable results if the correlations actually exist. As the SSI is defined under the assumption of independent parameters, methods for variance-based sensitivity analysis of systems with correlated random variables were introduced in the literature, such as [24,29,30,54,62]. The covariance decomposition-based sensitivity analysis (CoDSA) proposed by Li et al. [24] and further explained by Sudret and Caniou [50] is a promising tool for dealing with problems in the presence of correlations. In contrast to the SSI, MISA is not based on the assumption of independent parameters, and thus can be directly extended to problems with correlated parameters [4].

The SSI, CoDSA and MISA are commonly computed by using Monte Carlo simulation [4,24,47]. However, Monte Carlo simulations require a large number of deterministic simulations and become computationally expensive especially for large-scale problems. In order to confront the problem, a surrogate model is used to substitute the computationally expensive model. To this end, a polynomial chaos expansion (PCE) surrogate model is implemented [11]. The PCE model is estimated by using least angle regression (LAR[13];). The sensitivity measures are computed through the coefficients of the PCE model or rapid PCE model evaluations [50].

In this article, we answer two questions: (1) How does the performance of the (co)variance-based SA compare with that of the MISA in chemical processes? (2) How do the correlations among parameters impact the parameter sensitivities? To this end, we first provide a deep insight into the different implemented methods for GSA; i.e., the SSI, CoDSA and MISA approach. Second, we analyze the performance of the proposed methods with an application to a continuous-flow reactor in the absence and presence of parameter correlations by following the framework shown in Fig. 2.

The remainder of this paper is organized as follows. Section 2 presents the method for describing and sampling independent/correlated random variables. Section 3 describes the mathematical formulation and properties of the SSI, CoDSA and MISA. Section 4 presents the PCE, LAR and computational framework of these sensitivity measures. In Section 5, the application to a continuous-flow reactor is illustrated and discussed. The conclusion is given in Section 6.

2. Independent and correlated random model parameters

Before starting with the methods for sensitivity analysis, we briefly explain some definitions for multivariate distribution and the difference between independent and correlated model parameters.

In this article, the uncertainties are assigned to the model parameters and described with a specific type of probability distribution, for example, normal and uniform distribution. Note that the model parameters could also be design variables, which depends on the type of the problems we have.The model response Y is represented as the function G:

$$Y = G(\mathbf{X}) = G(X_1, X_2, ..., X_n), \tag{1}$$

where $X_1, X_2, ..., X_n$ are n random model parameters. The probability of the random parameters on their entire domain are given by an n-dimensional joint probability distribution ($f_{\mathbf{X}}(\mathbf{x})$), while the probability of X_i without reference to the value of the other parameters and with reference to the values of the other parameters is given by the marginal distribution ($f_{X_i}(x_i)$) and the conditional distribution ($f_{X_i|X_{\sim i}=x_{\sim i}}(x_i)$), respectively. Samples for model parameters could be drawn from one of these distributions based on the purpose. For example, a sample group with a constant value for X_i is generated from the conditional distribution $f_{X_{\sim i}|X_i=x_i}(x_{\sim i})$. Note that choosing the right distribution is extremely important because the conditional and marginal distributions are completely different in the case of correlated model parameters.

For independent model parameters, the conditional distribution for X_i is not affected by the values of other parameters and is defined by its marginal distribution. Their joint distribution is simplified to the product of the marginal distributions of each parameter:

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1}(x_1) f_{X_2}(x_2) ... f_{X_n}(x_n) \tag{2}$$

Note that the marginal distributions for each parameters are always given in advance. An example of independent parameters is shown in Fig. 1 (D) which illustrates independent bivariate normal distribution.

For the correlated model parameters, the equivalence between the marginal and conditional distribution and the simplification of the joint

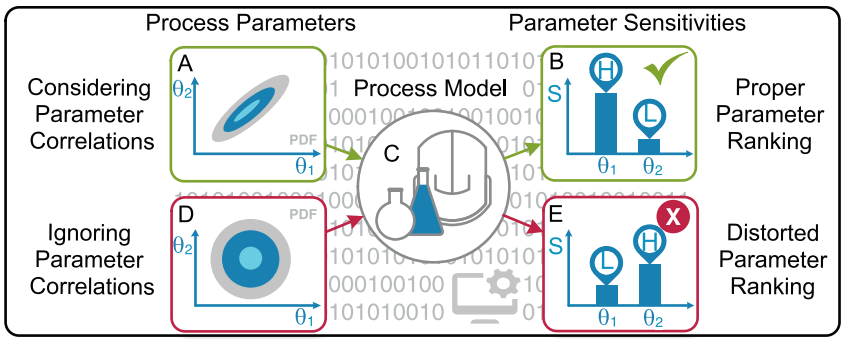


Fig. 1. Exemplary illustration of the parameter correlation effect on sensitivity analysis: A correlated bivariate probability density function of two process parameters **(A)** results into a proper parameter sensitivity ranking **(B)**. Parameter θ_1 has a high sensitivity (H), and the sensitivity of θ_2 is low (L). For the same process model **(C)** but ignoring parameter correlations **(D)** the sensitivity ranking is misleading **(E)**.

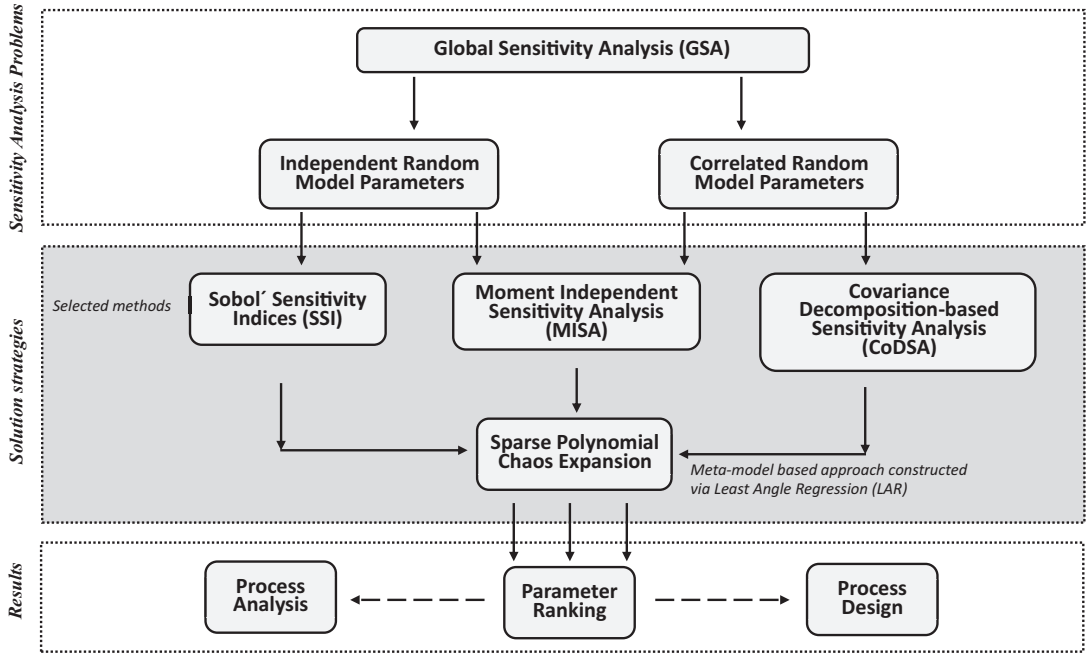


Fig. 2. Framework for global sensitivity analysis, upper rectangle gives problems, middle rectangle presents the methods to solve the sensitivity analysis problems, lower rectangle presents the results and their potential applications.

distribution in Eq. (2) no longer hold. In order to obtain the joint distribution for correlation model parameters, copula formalism [33] derived from Sklar's theorem provides the link between the joint and marginal distributions for all types of distributions no matter whether they are independent or not [45]. The copula formalism is written as follows:

$$f_{\mathbf{X}}(\mathbf{x}) = c(F_{X_1}(x_1),...,F_{X_n}(x_n))\cdot \prod_{i=1}^{n} f_{X_i}(x_i), \tag{3}$$

where $F_{X_i}(x_i)$ is the marginal cumulative density for random parameter X_i. Here, $c(F_{X_1}(x_1), ...,F_{X_n}(x_n))$ is the copula density which can be derived by transforming the copula function $C(F_{X_1}(x_1), ...,F_{X_n}(x_n))$:

$$c(F_{X_1}(x_1),...,F_{X_n}(x_n)) = \frac{\partial^n C(F_{X_1}(x_1),...,F_{X_n}(x_n))}{\partial F_{X_1}(X_1)...\partial F_{X_n}(x_n)} \tag{4}$$

One frequently used copula function is the Gaussian copula which is formulated as:

$$C(F_{X_1}(x_1),...,F_{X_n}(x_n);\boldsymbol{\rho}) = F_n(F^{-1}(F_{X_1}(x_1)),...,F^{-1}(F_{X_n}(x_n));\boldsymbol{\rho}) \tag{5}$$

Here, $F_n(\cdot)$ is the multivariate cumulative Gaussian distribution function with correlation matrix ρ, and $F^{-1}(\cdot)$ is the inverse standard Gaussian distribution function. The correlation matrix consists of Pearson's correlation coefficients (ρ_{ij}), which quantifies the correlation between X_i and X_j and is defined as:

$$\rho_{ij} = \frac{cov(X_i, X_j)}{\sigma_i \sigma_j}, \tag{6}$$

where *cov* is the covariance function, and σ_i is the standard deviation of X_i. The values of the correlation coefficient ρ are within the range −1 to 1, where 0 and 1(−1) represent independent and completely positive (negative) correlated variables, respectively. The Gaussian copula is available only for linear dependence which is shown in Fig. 1 (A). For variables with non-linear correlations, other copulas are required. Additional descriptions of other copulas are beyond the scope of this paper and refer to [33]. For chemical engineering applications, the correlation matrix is always used, and thus, the Gaussian copula is a proper option for this paper.

Samples of correlated and independent bivariate distributions are illustrated in Fig. 1 (A) and (D), respectively. Although the variables have the same marginal distributions, but the shapes of the sample spaces are completely different. The correlation between two parameters restricts the sampling space, which completely changes the information obtained from the sensitivity analysis.

3. Methods for global sensitivity analysis in the absence and presence of correlation among the model parameters

In this section, we briefly review the mathematical formulation and definition of the sensitivity measures for the methods used in the current article for sensitivity analysis, i.e., the SSI, CoDSA and MISA.

3.1. Sobol' Sensitivity indices (SSI) for independent model parameters

The SSI derived from the decomposition of the model output variance [46] is widely used for global sensitivity analysis of problems with independent model parameters.

For function $Y = G(\mathbf{X})$ of finite variance, it can be uniquely decomposed as follows:

$$G(\mathbf{X}) = G_0 + \sum_{i=1}^{n} G_i(X_i) + \sum_{1<i<j<n} G_{ij}(X_i, X_j) + \cdots + G_{12\cdots n}(X_1,...,X_n), \tag{7}$$

where the partial functions are defined as:

$$G_0 = E(Y), \tag{8}$$

$$G_{\mathbf{u}}(\mathbf{X_u}) = E_{\mathbf{X_{\sim u}}}(Y|\mathbf{X_u}) - \sum_{\substack{\mathbf{w}\subset\mathbf{u}\\ \mathbf{w}\neq\emptyset}} G_{\mathbf{w}} - G_0, \tag{9}$$

where $E(\cdot)$ denotes the expectation operation. The uniqueness of the function decomposition is ensured based upon the following properties of the partial functions:

$$\int G_{\mathbf{u}}(\mathbf{X_u}) f_{X_i}(x_i) dX_i = 0 \qquad i \in \mathbf{u}, \tag{10}$$

$$\int G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}})G_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}})f_{\mathbf{X}}(\mathbf{x})d\mathbf{X} = 0 \qquad \mathbf{u} \neq \mathbf{v} \tag{11}$$

$\mathbf{u}$ and $\mathbf{v}$ are different subsets of the full index set $\{1, ..., n\}$. Note that $G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}) = G_0$ if $\mathbf{u}$ is an empty set. $\mathbf{X}_{\mathbf{u}}$ includes the variables marked with the numbers in subset $\mathbf{u}$.

Based on the orthogonal property of the component function described in Eq. (11), the decomposition of the variance of function $G(X)$ can be deduced from Eq. (7) and is written as:

$$Var(Y) = \sum_{i=1}^{n} V_i + \sum_{1<i<j<n} V_{ij} + \cdots + V_{12\ldots n}, \tag{12}$$

where

$$V_{\mathbf{u}} = V_{\mathbf{X}_{\mathbf{u}}}(E_{\mathbf{X}_{\sim\mathbf{u}}}(Y|\mathbf{X}_{\mathbf{u}})) - \sum_{\substack{\mathbf{w}\subset\mathbf{u} \\ \mathbf{w}\neq\emptyset}} V_{\mathbf{w}}, \tag{13}$$

and $\sim\mathbf{u}$ denoting the complementary subset of $\mathbf{u}$. $Var(Y)$ is the variance of function $G(\mathbf{X})$. V_i, V_{ij}, ..., and $V_{123\ldots n}$ are partial variances which describe the effect of individual parameters or parameter interactions on $Var(Y)$.

Based on the normalization of Eq. (12), [46] introduced the SSI as follows:

$$S_i^{uc} = \frac{V_i}{Var(Y)}, \tag{14}$$

$$S_{ij}^{uc} = \frac{V_{ij}}{Var(Y)}, \tag{15}$$

...

$$S_{12\ldots,n}^{uc} = \frac{V_{12\ldots n}}{Var(Y)} \tag{16}$$

S_i^{uc} are called the first-order sensitivity indices and S_{ij}^{uc}, ..., $S_{12\ldots n}^{uc}$ are called interaction sensitivity indices. First-order sensitivity indices give the percentage of the total variance of Y due to the uncertainty in each input variable, whereas the interaction sensitivity indices measure the interaction among input variables. Moreover, the total sensitivity indices S_{Ti}^{uc} are introduced to describe the contribution of single variables X_i and their interactions with other model parameters, which is defined as [40]:

$$S_{Ti}^{uc} = \frac{V_i + V_{i,j} + \cdots + V_{1\ldots i\ldots n}}{Var(Y)} \tag{17}$$

Practically, instead of calculating $2^n - 1$ sensitivity indices, which are the total number of the first-order and interaction terms and increase dramatically with larger n, only the total sensitivity indices and the first-order sensitivity indices (with a number of $2n$) are calculated to save computational costs.

The SSI performs satisfactorily and reflects the model structure for functions with independent model parameters. However, a problem arises in implementing the SSI for functions with correlated model parameters as discussed in [34]. Therefore, other decomposition concept and new sensitivity indices are required for problems with correlated model parameters [24].

3.2. Covariance decomposition-based sensitivity analysis (CoDSA) for correlated model parameters

As we described in Section 2, the entire sampling space for correlated parameters is different from that for independent parameters. This might result in the shift of the mean and the variance, or more generally the entire distribution, of the model response as shown in Fig. 3. Therefore, it is necessary to calculate the sensitivities in the presence of parameter correlations.

One idea for dealing with correlated random parameters is using isoprobabilistic transformation concepts, such as the Rosenblatt transformation [37,53] and the Nataf transformation [21]. By utilizing those transformations, the original correlated variables are converted to new independent variables. For instance, this approach has been applied in [29,30]. Here, some new sensitivity measures are given regarding parameter correlations. Alternatively, as proposed by Li et al. [24], new sensitivity indices can be derived which are based on the covariance decomposition of function $G(\mathbf{X})$. As there is no need for an isoprobabilistic transformation step, the covariance decomposition approach is of primary interest in this work.

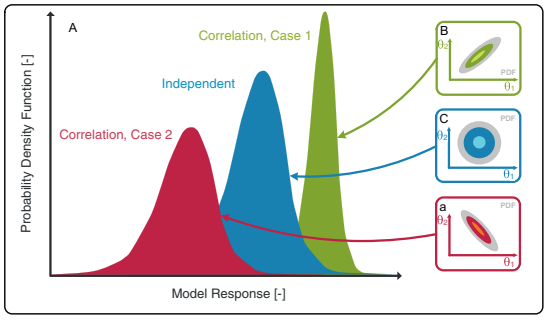


Fig. 3. Exemplary illustration of probability distributions of the model response (A) with independent (C) or correlated random parameters; i.e., Case 1 (B) and Case 2 (D) are two different situations decided by the structure of the model and dependency among the model parameters.

3.2.1. Covariance decomposition

Let us consider the second-order function $G(\mathbf{X}^{cor})$ again, but with correlated model parameters. As pointed out by Li et al. [25], the same functional decomposition of $G(\cdot)$ is still available, as the additional correlations among the model parameters do not affect the structure of the model. Therefore, similar functional relationships between the model parameters and the outputs can be derived as:

$$Y = G(\mathbf{X}^{cor}) = G_0 + \sum_{i=1}^{n} G_i(X_i^{cor}) + \sum_{1<i<j<n} G_{ij}(X_i^{cor}, X_j^{cor}) + \cdots + G_{123\cdots n}(X_1^{cor}, \cdots, X_n^{cor}), \tag{18}$$

$$G_0 = E(Y), \tag{19}$$

$$G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}^{cor}) = E_{\mathbf{X}_{\sim\mathbf{u}}}(Y) - \sum_{\mathbf{w}\subset\mathbf{u}} G_{\mathbf{w}} - \sum_{\substack{\mathbf{u}\neg\supseteq\mathbf{w}\subseteq\{1,2,\ldots,n\} \\ \mathbf{u}\cap\mathbf{w}\neq\emptyset}} E_{\mathbf{X}_{\sim\mathbf{u}}}[G_{\mathbf{w}}], \tag{20}$$

where $E_{\mathbf{X}_{\sim\mathbf{u}}}(\cdot)$ is the operator which calculates the (conditioned) mean value of the output with the marginal distribution of $\mathbf{X}_{\sim\mathbf{u}}$. After the decomposed functions are constructed, the correlations among the parameters should be considered for the sensitivity analysis. Note that $G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}^{cor})$ is not mutually orthogonal for the correlated model parameters, and thus, the variance decomposition of model output Y in Eq. (12) is no longer defined. Therefore, the covariance decomposition of $Var(Y)$ was introduced by Li et al. [24]:

$$\begin{aligned} Var(Y) &= Cov\left[\left(G_0 + \sum_{\substack{\mathbf{u}\subset\{1,2,\cdots,n\} \\ \mathbf{u}\neq\emptyset}} G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}^{cor})\right), G(\mathbf{X}^{cor})\right], \\ &= \sum_{\substack{\mathbf{u}\subset\{1,2,\cdots,k\} \\ \mathbf{u}\neq\emptyset}} Cov[G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}^{cor}), G(\mathbf{X}^{cor})], \end{aligned} \tag{21}$$

where $Var(Y)$ is exactly partitioned by $Cov[G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}^{cor}), G(\mathbf{X}^{cor})]$ which describes the total contribution of component function $G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}^{cor})$ to the total variance of output $Var(Y)$. Further decomposition of $Cov[G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}^{cor}), G(\mathbf{X}^{cor})]$ can be used to separate the total contribution into two parts similar to:

$$
\begin{aligned}
&Cov\left[G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}^{cor}),\, G(\mathbf{X}^{cor})\right] \\
&= Cov\left[G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}^{cor}),\, G_0 + \sum_{\substack{\mathbf{u}\subset\{1,2,\cdots,k\} \\ \mathbf{u}\neq\varnothing}} G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}^{cor})\right], \\
&= Var\left[G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}^{cor})\right] + \sum_{\mathbf{v}\neq\mathbf{u}} Cov\left[G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}^{cor}),\, G_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}}^{cor})\right]
\end{aligned} \tag{22}
$$

$Var\left[G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}^{cor})\right]$ denotes the contribution of the component function to $Var(Y)$, which depends on the function itself (structure) and the marginal pdf $f_{\mathbf{X}_{\mathbf{u}}}(\mathbf{X}_{\mathbf{u}})$ only. $\sum_{\mathbf{v}\neq\mathbf{u}} Cov\left[G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}^{cor}),\, G_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}}^{cor})\right]$ denotes the contribution due to the correlation among the model parameters.

Obviously, the covariance decomposition has a completely different structure compared to the variance decomposition for the independent case. The major difference between them is that the component contribution of the covariance decomposition (both covariance parts of Eqs. (21) and (22)) could be negative, which can never happen in variance-based decomposition. Due to the participation of correlations, new sensitivity indices are required for the sensitivity analysis.

3.2.2. Covariance-based sensitivity indices

Li et al.[25] introduced three new sensitivity indices for CoDSA, which are obtained by renormalizing Eqs. (21) and (22) with $Var(Y)$ and formulated as follows:

$$S_u^{cov} = Cov\left[G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}^{cor}),\, G(\mathbf{X}^{cor})\right]/Var(Y), \tag{23}$$

$$S_u^{U} = Var\left[G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}^{cor})\right]/Var(Y), \tag{24}$$

$$S_u^{C} = \sum_{\mathbf{v}\neq\mathbf{u}} Cov\left[G_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}^{cor}),\, G_{\mathbf{v}}(\mathbf{X}_{\mathbf{v}}^{cor})\right]/Var(Y) \tag{25}$$

The new indices are called total covariance-based sensitivity indices S_u^{cov}, structural sensitivity indices S_u^{U}, and correlative sensitivity indices S_u^{C} [25]. The structural and correlative sensitivity indices indicate, in turn, the sensitivity of model structure and correlations among the model parameters on model output Y. S_u^{cov} is the sum of the two sensitivity indices as shown in Eq. (26) and presents the total effect on model output Y:

$$S_u^{cov} = S_u^{U} + S_u^{C} \tag{26}$$

Note that the sum of S_u^{cov} for all possible subset $\mathbf{u}$ is equal to 1 no matter whether the individual value of S_u^{cov} is positive, negative or larger than 1. The three sensitivity indices could be the sensitivities for individual parameters (first-order) or interactions among two or more variables depending on subset $\mathbf{u}$. The total sensitivity indices can be directly extended for CoDSA similar to Eq. (17), and thus, we have the total covariance-based total sensitivity indices $S_{T_i}^{cov}$, structural total sensitivity indices $S_{T_i}^{U}$, and correlative total sensitivity indices $S_{T_i}^{C}$. The structure of all the indices used for CoDSA is illustrated in Fig. 4, where the structural sensitivity indices are listed on the left hand side and the correlative sensitivity are listed on the right hand side.

As mentioned previously, the assumption of mutual orthogonality of the component functions in Eq. (18) fails for correlated parameters. Thus, the decomposition in Eq. (18) is not unique and depends on the applied method, respectively. This may affect the sensitivity results derived from the covariance decomposition. Alternatively, [22] proposed a general and unique decomposition for functions with correlated parameters based on the relaxed vanishing condition, where the component functions are hierarchically orthogonal. Such kind of decomposition is more rigorous but requires sophisticated construction methods. The interested reader is referred to [22,23] and references therein.

3.3. Moment-independent sensitivity analysis (MISA)

In the previous subsections, we presented the SSI and CoDSA which are defined based on the (co)variance of the model output. However, using a single statistical moment for SA may lead to an apparent loss of information from other statistical moments as skewness and kurtosis [4]. Therefore, MISA, which is derived directly from the entire distribution of the model output, is introduced and used for sensitivity studies [4]. MISA compares the difference between probability distribution $f_Y(y)$ of output Y and conditional probability distributions $f_{Y|X_i}(y)$ of output Y to calculate the parameter sensitivities [4], which is mathematically expressed as:

$$s(X_i) = \int \left| f_Y(y) - f_{Y|X_i}(y) \right| dy \tag{27}$$

$s(X_i)$ is also called the shift function, and the average of the shift function on the entire distribution of X_i is then given by:

$$E_{X_i}[s(X_i)] = \int f_{X_i}(x_i) \left[\int \left| f_Y(y) - f_{Y|X_i}(y) \right| dy \right] dx_i, \tag{28}$$

where $f_{X_i}(x_i)$ is the marginal density of model parameter X_i. Based on Eq. (28), [4] proposed a new indicator δ_i for global sensitivity analysis, which is defined as follows:

$$\delta_i = \frac{1}{2} E_{X_i}[s(X_i)] \tag{29}$$

The indicator can also be directly extended to a group of parameters equal to:

$$E_{\mathbf{X}_{\mathbf{u}}}[s(\mathbf{X}_{\mathbf{u}})] = \int f_{\mathbf{X}_{\mathbf{u}}}(\mathbf{x}_{\mathbf{u}}) \left[\int \left| f_Y(y) - f_{Y|\mathbf{X}_{\mathbf{u}}}(y) \right| dy \right] d\mathbf{x}_{\mathbf{u}}, \tag{30}$$

$$\delta_{\mathbf{u}} = \frac{1}{2} E_{\mathbf{X}_{\mathbf{u}}}[s(\mathbf{X}_{\mathbf{u}})], \tag{31}$$

in which $\mathbf{u}$ is a vector that includes the index of the parameters part of the group, and $f_{\mathbf{X}_{\mathbf{u}}}(\mathbf{x}_{\mathbf{u}})$ is the marginal distribution of the parameter group. Borgonovo [4] also included 5 properties for the indicator: (1) The δ_i ($\delta_{\mathbf{u}}$) varies in the range [0,1] and (2) it indicates independence between X_i (group $\mathbf{X}_{\mathbf{u}}$) and output Y if $\delta_i(\delta_{\mathbf{u}}) = 0$. (3) The indicator of all model parameters ($\delta_{1,2,\ldots,n}$) equals unity, which is evident as the uncertainty of the output vanishes once the uncertainties of all model parameters are eliminated. (4) δ_{ij} for parameters X_i and X_j is bounded as

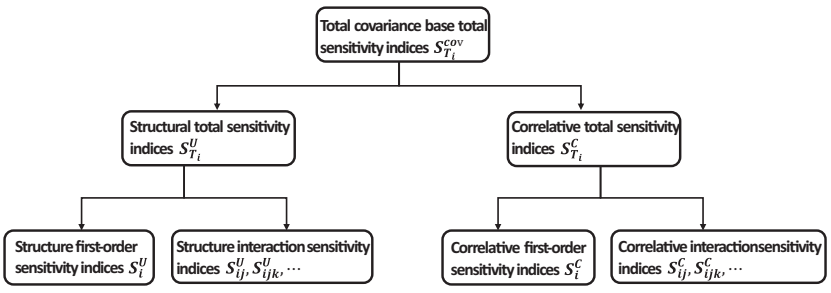


Fig. 4. Topology of the covariance-based sensitivity indices.

$\delta_i \le \delta_{ij} \le \delta_i + \delta_{i|j}$, in which $\delta_{i|j}$ is the conditional δ. (5) The left equal holds if the output is independent of X_j. The last two properties give the boundary for the parameter sensitivity indicator. These properties make the sensitivity indicator more representative and comparable. Further descriptions and proofs are found in [4]. The method is also available for problems with correlated model parameters, as the assumption of independent parameters is not required for its definition.

In this section, we reviewed three different methods for SA. However, the proposed concepts require the calculation of statistical values for the model output, which results in high-order numerical integration problems. Therefore, highly efficient methods for uncertainty quantification are required to lower the computational burden of applying SA to complex chemical processes.

4. Estimation of sensitivity measures using polynomial chaos expansion

Numerical techniques are the standard in determining parameter sensitivities because analytical solutions are not feasible, especially for complex chemical processes [11,40]. The main groups of the numerical methods for SA are as follows:

(i) Monte Carlo simulations or more efficient quasi-Monte Carlo simulations. The implementation of these methods is straightforward but requires an extensive evaluation of the original model.
(ii) The meta-model-based approach, e.g., polynomial meta-models, kriging meta-models and neural network meta-models. In general, the meta-model is used as a substitution of the original model to ease the computational burden, but the estimation of the meta-model requires some effort.

PCE is used in this article as PCE is tailored for SA [11]. The basics of PCE and an efficient implementation strategy are introduced in the following.

4.1. Polynomial chaos expansion

PCE was introduced in the 1930s by Wiener [58] and applied to the engineering field for solving stochastic finite element problems by Ghanem and Spanos [15]. It was available only for variables with standard Gaussian distribution at the beginning and generalized to other types of statistical distributions, e.g., uniform and beta distribution, by Xiu and Karniadakis [61].

In PCE, a function of finite variance $G(\boldsymbol{\xi}) \in L^2(\Omega, \mathscr{F}, \mathscr{P})$, can be represented as [58]:

$$G(\boldsymbol{\xi}) = a_0\Gamma_0 + \sum_{i_1=1}^{n_\xi} a_{i_1}\Gamma_1(\xi_{i_1}) + \sum_{i_1=1}^{n_\xi}\sum_{i_2=1}^{i_1} a_{i_1,i_2}\Gamma_2(\xi_{i_1},\xi_{i_2}) + \cdots + \sum_{i_1=1}^{n_\xi}\sum_{i_2=1}^{i_1}\cdots\sum_{i_p=1}^{i_{p-1}} a_{i_1,i_2,\cdots,i_p}\Gamma_p\left(\xi_{i_1},\xi_{i_2},\cdots,\xi_{i_p}\right) + \cdots, \tag{32}$$

where Γ_p is the pth order multivariate polynomial for $\boldsymbol{\xi} = \{\xi_i\}_{i=1}^{n_\xi}$. $\boldsymbol{\xi}$ are independent random variables with standard normal distribution. $a_{i_1\cdots,i_P}$ are the coefficients which quantify the relation between the corresponding polynomial Γ_P and the function $G(\boldsymbol{\xi})$. PCE can also be directly written as:

$$G(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} \alpha_k \Psi_k(\boldsymbol{\xi}), \tag{33}$$

where the $\{\Psi_k(\boldsymbol{\xi})\}_{k=0}^{\infty}$ and $\{a_k\}_{k=0}^{\infty}$ are multivariate polynomials and corresponding coefficients of infinite degree, respectively. k is the number of the polynomials. The multivariate polynomials $\Psi_k(\boldsymbol{\xi})$ are constructed by the product of the univariate polynomials [61]:

$$\Psi_k(\boldsymbol{\xi}) = \Phi_1^{k_1}(\xi_1)\Phi_2^{k_2}(\xi_2)\cdots\Phi_{n_\xi}^{k_{n_\xi}}\left(\xi_{n_\xi}\right), \tag{34}$$

where nonnegative integer $\{k_i\}_{i=1}^{n_\xi}$ indicates the individual order of the univariate polynomials $\{\Phi_i^{k_i}(\xi_i)\}_{i=1}^{n_\xi}$. The order of the multivariate polynomials $\Psi_k(\boldsymbol{\xi})$ is the sum according to $p = k_1 + k_2 + \cdots + k_{n_\xi}$. The multivariate and univariate polynomials are orthogonal with respect to their corresponding stochastic measure:

$$\int \Phi_i^m(\xi_i)\Phi_i^n(\xi_i) f(\xi_i) d\xi_i = r_m \delta_{mn}, \tag{35}$$

$$\int \Psi_m(\boldsymbol{\xi})\Psi_n(\boldsymbol{\xi}) f(\boldsymbol{\xi}) d\boldsymbol{\xi} = \gamma_m \delta_{mn}, \tag{36}$$

$$f(\boldsymbol{\xi}) = f(\xi_1) f(\xi_2)\cdots f\left(\xi_{n_\xi}\right) d\xi_1 d\xi_2 \cdots d\xi_{n_\xi}, \tag{37}$$

where δ_{mn} is the Kronecker delta function which is 1 for identical values of m and n and 0 for the others. r_m and γ_m are the normalized constants of Φ_m and Ψ_m, respectively.

The type of univariate polynomials depends on the probability distributions of the random variables. Wiener [58] introduced Hermite polynomials for Gaussian random variables. Although the model can be used for other distributions through isoprobabilistic transformation, the efficiency and accuracy of the model are decreased considerably [60]. Alternatively, the Askey scheme was used to construct orthogonal polynomials for other types of statistical distribution, e.g., Jacobi polynomials for beta distribution, as shown in [61]. Subsequently, PCE has been further adapted to arbitrary distributions by Wan and Karniadakis [35,56,59]. The general formulation of PCE works only for independent variables, but we can also construct PCE for correlated variables [32]. However, in this work, we include directly the parameter correlations in SA rather than in PCE model.

Practically, instead of using a set of an infinite number of polynomials as in Eq. (33), a finite number of polynomials is retained to approximate the random variable [51], e.g., polynomials with a total order not exceeding p:

$$G(\boldsymbol{\xi}) \approx \sum_{k=0}^{P-1} \alpha_k \Psi_k(\boldsymbol{\xi}), \tag{38}$$

where P is the dimension of the polynomials basis depending on maximum order p and n_ξ:

$$P = \binom{n_\xi + p}{p} = \frac{(n_\xi + p)!}{n_\xi! p!} \tag{39}$$

The dimension of the input variable (n_ξ) is fixed for certain problems. The optimal order p is the minimum value we need to guarantee the target accuracy.

4.2. PCE coefficient calculation

Estimation of the coefficients $\{\alpha_k\}_{k=0}^{P-1}$ plays an essential role in PCE-based calculations as the stochastic properties of PCE are mainly characterized by them. Several methods are available for computing the coefficients, and they are classified into two groups: intrusive methods and non-intrusive methods. Intrusive methods, such as Galerkin projection, have optimal accuracy but require adaptation of the numerical model, which might be challenging to implement for chemical processes [60]. In contrast, non-intrusive methods require only model evaluations for some realizations and can be applied for models of any complexity. The number of required model evaluations, however, might restrict the implementation, too. Alternatively, [3] proposed an adaptive procedure to estimate the PCE model by using the sparsity of its coefficients. The procedure was developed from Least Angle Regression (LAR) [3,13] and summarized in Algorithm 1.

Here, let us consider a sample set $\mathscr{S} = \{\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^N\}$ generated for random variables $\boldsymbol{\xi}$. $\mathbf{Y} = (G(\boldsymbol{\xi}^1), \ldots, G(\boldsymbol{\xi}^N))$ is the vector of the model

Initialization

1: Select the truncated scheme
2: Compute $\mathbf{Y}$ and $\{\Psi_k(\boldsymbol{\xi})\}_{k=0}^{P-1}$ for $\mathcal{S}$
Estimation of polynomial coefficients
3: **for** $p \leftarrow 1, p_{\max}$ **do**
4: Set $\mathbf{a} = \{\alpha_0, \ldots, \alpha_{P-1}\} = \mathbf{0}$, $\mathbf{R} = \mathbf{Y}$, active set = { }, basic set = $\{\Psi_k(\boldsymbol{\xi})\}_{k=0}^{P-1}$, $m = 0$
5: **while** $m \leq \min(N, P)$ **do**
6: $k^* = \underset{\Psi_k(\boldsymbol{\xi}) \in \text{basic set}}{\text{argmax}} \; |Corr(\mathbf{R}, \Psi_k(\boldsymbol{\xi})|$
7: Move basis polynomial $\Psi_{k^*}(\boldsymbol{\xi})$ from the basic set to the active set
8: Calculate the correction term $\boldsymbol{\Delta}$ [13]
9: Update the polynomial coefficients, $\mathbf{a} = \mathbf{a} + \Delta$
10: Update the residual, $\mathbf{R} = \mathbf{Y} - \sum_{k=0}^{P-1} \alpha_k \Psi_k(\boldsymbol{\xi})$
11: Recalculate the coefficients for the active set with
12: ordinary least squares regression
13: Get mean approximation error $\overline{\varepsilon}_m^{(p)}$ via the cross-validation procedure
14: $m = m + 1$
15: **end while**
16: Store $\overline{\varepsilon}^* = \min(\overline{\varepsilon}_1^{(p)}, \ldots, \overline{\varepsilon}_m^{(p)})$ and the corresponding optimal active set
17: **Stop if** either $\overline{\varepsilon}^*$ satisfies the target accuracy or increases for the last two iterations
18: **end for**
19: Estimate the relevant polynomial coefficients for the last optimal active set via ordinary least squares regression

Algorithm 1. Computing adaptive sparse PCE with LAR.

evaluations associated with the sample set $\mathscr{S}$. The function $G(\boldsymbol{\xi})$ is approximated by $\sum_{k=0}^{P-1} \alpha_k \Psi_k(\boldsymbol{\xi})$, where P is the number of polynomials of the truncated scheme (Eq. (39)). Note that other truncated schemes can be selected based on the information we have for the model structure [3]. The algorithm is initialized by function evaluations $\mathbf{Y}$ and of polynomials $\{\Psi_k(\boldsymbol{\xi})\}_{k=0}^{P-1}$ from the truncated scheme, which are evaluated for sample set $\mathscr{S}$. Note that the size of the sample set can be enlarged when the accuracy of the PCE is not satisfied. For each iteration, the size of the multivariate polynomials is enlarged while the maximum order of the polynomials increases from 1 to $p_{\max}$, where $p_{\max}$ is the maximum order of the multivariate polynomials allowed for approximation. The most correlated polynomial is selected and moved to the active set at step 6. The coefficients for the polynomials of the active set are adapted in an optimal direction in step 9, using the correction term Δ [13]. All the possible active sets are validated, and the optimal one is selected by using cross validation in steps 11 and 12. Finally, the coefficients for the optimal PCE (step 18) are estimated through ordinary least squares regression. Note that the size of the active set, i.e., the number of multivariate polynomials used for the final estimation, is typically much smaller than the size of the full PCE model. This leads to a considerable reduction in the number of model evaluations. In principle, the resulting sparsity of the PCE model can be quantified by

$$Sparsity = \frac{\mathcal{A}(\text{Active set})}{\mathcal{A}(\text{Full set})}, \tag{40}$$

where $\mathcal{A}$ means the cardinality of the set.

4.3. Computation of sensitivity indices using PCE

Suppose we have a function $Y = G(\boldsymbol{X})$, where $\boldsymbol{X}$ is the n_X-dimensional random model parameter vector with given distributions. With the proposed procedures in the last section, the function $Y = G(\boldsymbol{X})$ is represented by the PCE truncated at order p:

$$Y = G(\boldsymbol{X}) \approx \sum_{k=0}^{P-1} \alpha_k \Psi_k(\boldsymbol{X}) \tag{41}$$

For the sake of simplicity, we define index $\mathbf{k} = \{k_1, k_2, \ldots, k_{n_X}\}$, where k_i is the order of univariate polynomials for individual input variable i. A is a set of all possible index $\mathbf{k}$ truncated at order p:

$$A = \{\mathbf{k} \in \mathbb{N}^{n_X} | k_1 + k_2 + \cdots + k_{n_\xi} \leq p\} \tag{42}$$

Therefore, the PCE can be written compactly as:

$$Y = G(\boldsymbol{X}) \approx \sum_{\mathbf{k} \in A} \alpha_{\mathbf{k}} \Psi_{\mathbf{k}}(\boldsymbol{X}) \tag{43}$$

4.3.1. Computation of the SSI using PCE

To calculate the SSI, function $G(\boldsymbol{X})$ is decomposed as shown in Eq. (7). Due to the orthogonality of the basis polynomials, the component functions $G_{\mathbf{u}}$ in the Sobol' decomposition in Eq. (7) are approximated by:

$$G_{\mathbf{u}}(\boldsymbol{X}_{\mathbf{u}}) \approx \sum_{\mathbf{k} \in A_{\mathbf{u}}} \alpha_{\mathbf{k}} \Psi_{\mathbf{k}}(\boldsymbol{X}), \tag{44}$$

where $A_{\mathbf{u}} \subset A$ and is defined as:

$$A_{\mathbf{u}} = \{\mathbf{k} \in A | k_i \neq 0 \text{ if and only if } i \in \mathbf{u}\} \tag{45}$$

From the definition, it is clear that the PCE in Eq. (44) is only a function of the variables included in $\boldsymbol{X}_{\boldsymbol{u}}$ because the order of the univariate polynomials of the other variables is zero. The statistical property of function $G(\boldsymbol{X})$, especially the mean and the variance, can be directly obtained from the coefficients of their PCE coefficients as:

$$E_{Y,A} = \alpha_0, \tag{46}$$

$$V_{Y,A} = \sum_{\mathbf{k} \in \mathbf{A}, \mathbf{k} \neq \mathbf{0}} \alpha_{\mathbf{k}}^2 \tag{47}$$

The SSI (Eq. (14)) can easily be derived from the representation above:

$$S_{\mathbf{u}}^{uc} = \frac{1}{V_{Y,A}} \sum_{\mathbf{k} \in A_{\mathbf{u}}} \alpha_{\mathbf{k}}^2, \tag{48}$$

where $S_{\mathbf{u}}^{uc}$ could be 1) a first-order sensitivity if $\mathbf{u}$ contains only one element or 2) an interaction sensitivity if it contains more than one. The total sensitivity indices are then given by:

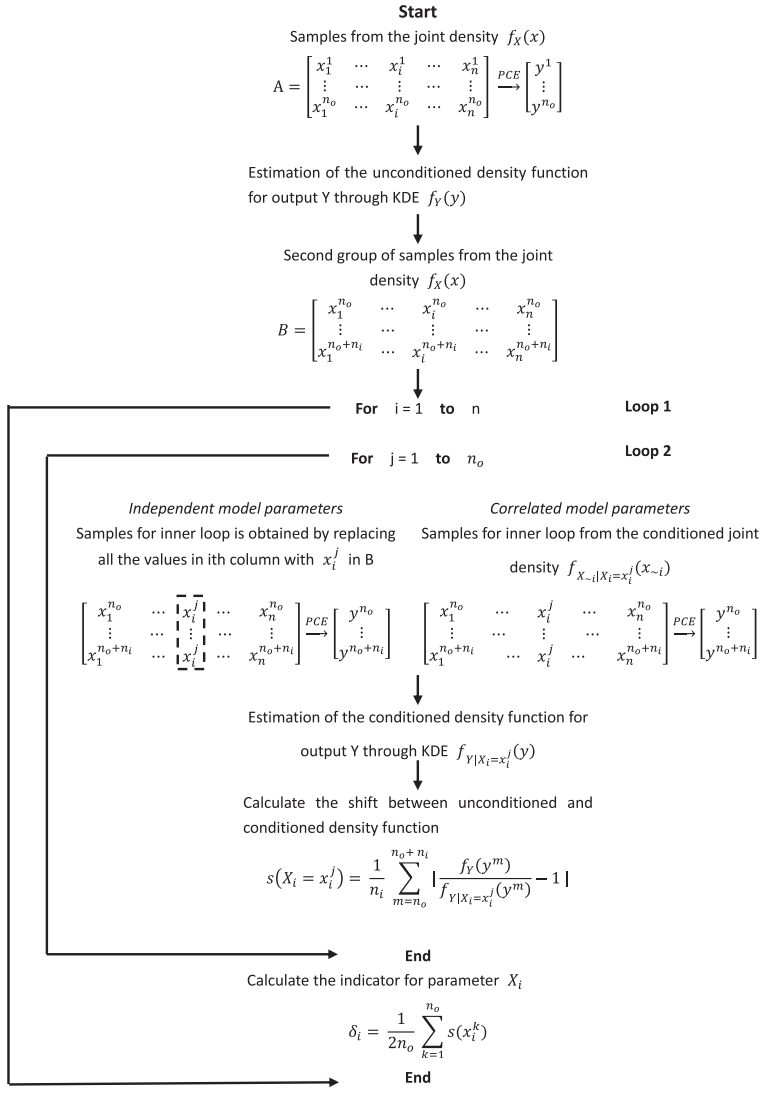


Fig. 5. The framework of the double-loop method for the calculation of indicator δ [57].

$$S_{T_i}^{uc} = \frac{1}{V_{Y,A}} \sum_{\mathbf{k} \in \overline{A_i}} \alpha_{\mathbf{k}}^2, \tag{49}$$

where $\overline{A_i} \subset A$ and includes all the basis polynomials related to variable ξ_i:

$$\overline{A_i} = \{\mathbf{k} \in A | k_i \neq 0\} \tag{50}$$

4.3.2. Computation of CoDSA using PCE

The construction of the sensitivity indices for function $Y = G(\boldsymbol{X})$ with correlated inputs is presented in Section 3.2.2, where three new sensitivity indices S_u^{cov}, S_u^U and S_u^C are introduced. As we explained in Section 3.2.2, a unique decomposition where the component functions are hierarchically orthogonal is not easy to construct. Therefore, we use the function decomposition approach, i.e., the PCE model, to compute the sensitivity indices. The exclusive use of PCE models also makes the computation more comparable for the three GSA techniques. The PCE approximation, as shown in Eqs. (43) and (44), is constructed as if the random variables are independent. However, the statistical moments in covariance decomposition cannot be derived directly from the coefficients of PCE because the orthogonality between the basis polynomials of PCE does not exist for the correlated variables. For this reason, Monte Carlo simulation are used to estimate the mean, variance and covariance of function $G(\boldsymbol{X})$ and component function $G_{\mathbf{u}}(\boldsymbol{X}_{\boldsymbol{u}})$ with samples drawn from correlated distributions:

$$E(Y) = \frac{1}{N} \sum_{i=1}^{N} \sum_{\mathbf{k} \in A} \alpha_{\mathbf{k}} \Psi_{\mathbf{k}}(\boldsymbol{X}^i), \tag{51}$$

$$Var(Y) = \frac{1}{N-1} \sum_{i=1}^{N} \left(\sum_{\mathbf{k} \in A} \alpha_{\mathbf{k}} \Psi_{\mathbf{k}}(\boldsymbol{X}^i) - E(Y) \right)^2, \tag{52}$$

$$E(G_{\mathbf{u}}(\boldsymbol{\xi}_{\boldsymbol{u}})) = \frac{1}{N} \sum_{i=1}^{N} \sum_{\mathbf{k} \in A_{\mathbf{u}}} \alpha_{\mathbf{k}} \Psi_{\mathbf{k}}(\boldsymbol{X}_{\boldsymbol{u}}^i) \tag{53}$$

Here, $\boldsymbol{X}^i$ is one sample vector for the random model parameter vector, and n samples are drawn from the given distribution. Eqs. (54) and (55) represent an evaluation of the functions approximated by PCE with the sample $\boldsymbol{X}^i$:

$$G(\boldsymbol{X}^i) = \sum_{\mathbf{k} \in A} \alpha_{\mathbf{k}} \Psi_{\mathbf{k}}(\boldsymbol{X}^i), \tag{54}$$

$$G_{\mathbf{u}}(\boldsymbol{X}_{\mathbf{u}}^i) = \sum_{\mathbf{k} \in A_{\mathbf{u}}} \alpha_{\mathbf{k}} \Psi_{\mathbf{k}}(\boldsymbol{X}^i) \tag{55}$$

The covariance-based sensitivity indices computed from the Monte

Carlo estimation are formulated as:

$$S_u^{cov} = \frac{\sum_{i=1}^{N}(G_{\mathbf{u}}(\boldsymbol{X}_{\mathbf{u}}^i) - E(G_{\mathbf{u}}(\boldsymbol{X}_{\boldsymbol{u}})))(G(\boldsymbol{X}^i) - E(Y))}{(N-1)Var(Y)}, \tag{56}$$

$$S_u^{U} = \frac{\sum_{i=1}^{N}(G_{\mathbf{u}}(\boldsymbol{X}_{\mathbf{u}}^i) - E(G_{\mathbf{u}}(\boldsymbol{X}_{\boldsymbol{u}})))^2}{(N-1)Var(Y)}, \tag{57}$$

$$S_u^{C} = S_u^{cov} - S_u^{U} \tag{58}$$

We should note that the computational cost is low as all the evaluations are conducted on the PCE model. The computational effort for the Monte Carlo simulation on the PCE model is negligible compared to that for estimation of the coefficients for the PCE model.

4.4. Computation of MISA using PCE

According to the definition of MISA in Section 3.3, the density functions $f_Y(y)$ and $f_{Y|X_i}(y)$ for the model output are required for calculating the shift function and the indicator, and therefore, the kernel density estimator (KDE) is used to estimate the distributions [7]. Moreover, the double-loop Monte Carlo method illustrated in Fig. 5 is used to calculate the indicator δ [57].

The double-loop method starts with two groups of samples generated from the distribution of the random model parameters. The unconditioned density function $f_Y(y)$ is estimated from the model evaluations associated with the samples in A. For loop 2, the conditioned density function $f_{Y|X_i}(y)$ is estimated by n_i samples, and the shift function $s(X_i)$ is calculated at a given value x_i^j. Note that the samples are different for the independent and correlated model parameters. For the independent case, the samples are generated directly by replacing the ith column of sample B with the (i, j)th element x_{ij} in sample A. However, this does not work for the correlated case as the conditioned density $f_{X_{\sim i}|X_i=x_i^j}(x_{\sim i})$ is different for different values of X_i. Therefore, the samples should be updated for each iteration according to the conditioned density with parameter X_i specified to the value x_i^j in sample A. In loop 1, loop 2 is repeated for different values of X_i from sample A, and the indicator δ_i is calculated with the values for the shift function. Note that the calculation formula for shift function $s(X_i)$ and δ_i is the Monte Carlo estimation of Eqs. (27) and (29). The total number (n_t) of model evaluations for the double-loop Monte Carlo method is $(n \times n_i + 1) \times n_o$, where n, n_o and n_i are the number of model parameters, the size of samples A and B, respectively. The number n_t could be prohibitively high as the KDE requires a large sample size to ensure the accuracy of the estimated density function. Thus, the PCE model derived above is used for the model evaluations to ensure low computational costs.

5. Case study: A continuous synthesis of an API–scaffold

5.1. Problem statement

In this study, we consider a model of a continuous-flow reactor processing the synthesis of aminopyrimidine as an API–scaffold [36]. The mechanism of the reactions is described as follow:

$$A + B \xrightarrow{r_1} C \tag{59}$$

$$A + B \xrightarrow{r_2} D \tag{60}$$

$$C + B \xrightarrow{r_3} E \tag{61}$$

$$D + B \xrightarrow{r_4} E \tag{62}$$

Eq. (59) to Eq. (62) describe the nucleophilic aromatic substitution reactions (S_NAr) of 2,4-dichloropyrimidine (A) and morpholine (B) in ethanol which produce the desired product 2-substituted aminopyrimidine (D), the less-desired product 4-substituted (C) and side product 2,4-substituted (E). A complete description of the reactions and their potential application in the pharmaceutical field can be found in [36]. The governing equations of the continuous-flow reactor are formulated as follows:

$$\frac{dC_A}{dt} = -k_1 C_A C_B - k_2 C_A C_B \tag{63}$$

$$\frac{dC_B}{dt} = -k_1 C_A C_B - k_2 C_A C_B - k_3 C_B C_C - k_4 C_B C_D \tag{64}$$

$$\frac{dC_C}{dt} = k_1 C_A C_B - k_3 C_B C_C \tag{65}$$

$$\frac{dC_D}{dt} = k_2 C_A C_B - k_4 C_B C_D \tag{66}$$

$$\frac{dC_E}{dt} = k_3 C_B C_C + k_4 C_B C_D, \tag{67}$$

where

$$k_i = A_i exp\left(-\frac{E_{Ai}}{RT}\right), \quad i \in \{1, 2, 3, 4\} \tag{68}$$

in which C_j with $j \in \{A, B, C, D, E\}$ are the concentrations, k_i are the reaction constants, A_i and E_{Ai} are pre-exponential factors and activation energies, and r_i are the reaction rates determined by the corresponding reaction constant and concentration of reactants. According to [36] an isothermal reactor is assumed, i.e., the temperature T is constant along the reactor. Residence time t is defined as the position in the tubular reactor divided by the flow rate of the substance on the supposition that an ideal fluid is inside the reactor. The values for the initial conditions and parameters are listed in Table 1, where t_{end} is the final residence time which decides the length of the tubular reactor. The kinetic parameters A_i and E_{Ai} estimated from the experiments are not accurate and their uncertainties are characterized by normal distributions as shown in Table 1. For the first part of the sensitivity analysis, the eight parameters are assumed to be independent. However, the data provided by Reizman and Jensen [36] reveal strong correlations among the parameters measured by the correlation matrix in Table 2. This correlation matrix is used in the second part of the sensitivity analysis. The model output we are interested in is the final concentration of product D (C_{Df}). Therefore, we construct the surrogate model for the eight kinetic parameters and C_{Df} before starting the sensitivity analysis. The PCE model is constructed and estimated in UQLAB© [31], software for UQ and surrogate modeling. Sensitivity analysis performed later are conducted in MATLAB® where function *ode15s* is used to solve the dynamic equations of the reactor.

Table 1
Parameters and uncertainties for the continuous-flow reactor [36].

Parameters		Nominal value	Uncertainty
C_{A0}(M)		0.150	–
C_{B0}(M)		0.375	–
C_{C0}(M)		0	–
C_{D0}(M)		0	–
C_{E0}(M)		0	–
T(K)		373.15	–
t_{end}(s)		1200	–
R(J/mol·k)		8.314	–
$\log_{10}(A_1)(M^{-1}s^{-1})$	(θ_1)	3.4	$\mathcal{N}(3.4\ ,\ 0.1)$
E_{A1}(kJ/mol)	(θ_2)	27.0	$\mathcal{N}(27.0\ ,\ 0.6)$
$\log_{10}(A_2)(M^{-1}s^{-1})$	(θ_3)	3.5	$\mathcal{N}(3.5\ ,\ 0.1)$
E_{A2}(kJ/mol)	(θ_4)	32.1	$\mathcal{N}(32.1\ ,\ 0.6)$
$\log_{10}(A_3)(M^{-1}s^{-1})$	(θ_5)	4.9	$\mathcal{N}(4.9\ ,\ 0.2)$
E_{A3}(kJ/mol)	(θ_6)	60.0	$\mathcal{N}(60.0\ ,\ 1.6)$
$\log_{10}(A_4)(M^{-1}s^{-1})$	(θ_7)	3.0	$\mathcal{N}(3.0\ ,\ 0.2)$
E_{A4}(kJ/mol)	(θ_8)	45.0	$\mathcal{N}(45.0\ ,\ 1.7)$

Table 2
Correlation coefficients for the eight parameters from [36].

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
θ_1	1.000	0.997	0.976	0.968	−0.002	−0.003	0.000	0.000
θ_2	0.997	1.000	0.976	0.973	−0.003	−0.003	0.000	0.000
θ_3	0.976	0.976	1.000	0.997	−0.006	−0.006	0.000	0.000
θ_4	0.968	0.973	0.997	1.000	−0.007	−0.007	0.000	0.000
θ_5	−0.002	−0.003	−0.006	−0.007	1.000	1.000	−0.008	−0.008
θ_6	−0.003	−0.003	−0.006	−0.007	1.000	1.000	−0.008	−0.008
θ_7	0.000	0.000	0.000	0.000	−0.008	−0.008	1.000	1.000
θ_8	0.000	0.000	0.000	0.000	−0.008	−0.008	1.000	1.000

Table 3
PCE model settings and characteristics.

Number of random inputs	8
Polynomial basis	Hermite
Maximum order of polynomials	3
Number of model evaluations	200
Estimation error	0.001
Sparsity	40%

5.2. Construction of the PCE model

The PCE model for C_{Df} (model output) and eight kinetic parameters (inputs) is constructed based on the probability distribution assigned to the parameters. Table 3 lists the information for the PCE model. According to Table 3, the coefficients of the PCE model are estimated with 200 random evaluations of the original model by using Algorithm 1, and only 40% of the full basis which have maximum order of 3 are activated here. Please note that the maximum order of the polynomials is determined by the desired estimation error and the complexity of the reactor model. The accuracy of the PCE model is indicated by the estimation error in Table 3, and further analyzed by comparing with the results from direct Monte Carlo simulations in Fig. 6. The probability density function estimated by the PCE model with 200 model evaluations is as good as the one from Monte Carlo simulations with 10,000 model evaluations but much better than the one from Monte Carlo simulations with 200 model evaluations; i.e., using the original process model given in Eqs. (63) to (67). Based on the PCE model, we can then calculate different sensitivity measures in the absence and presence of correlations as in the following.

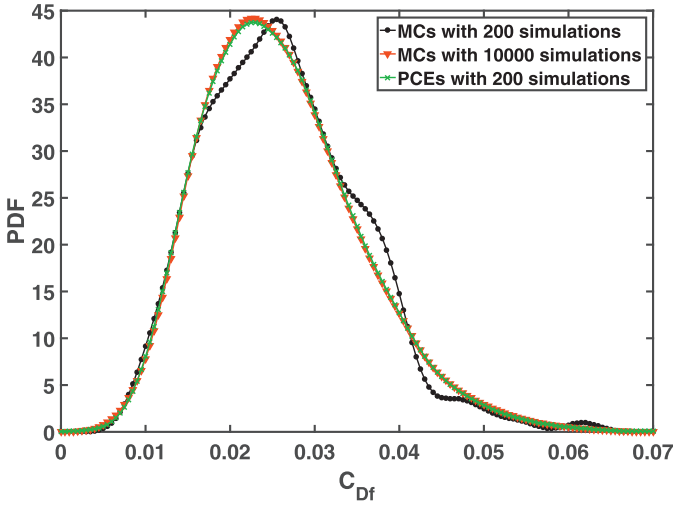


Fig. 6. Probability density function of component D (C_{Df}) evaluated with 200 and 10,000 Monte Carlo simulations and with PCE from 200 simulations. KDE is used to estimate the probability density functions with the model evaluations.

5.3. Sensitivity measures in the absence of correlation

This section presents the results for the first part of the sensitivity analysis where the correlations among the parameters are neglected. The SSI and MISA for the independent parameters on C_{Df} are calculated by using the sparse PCE model.

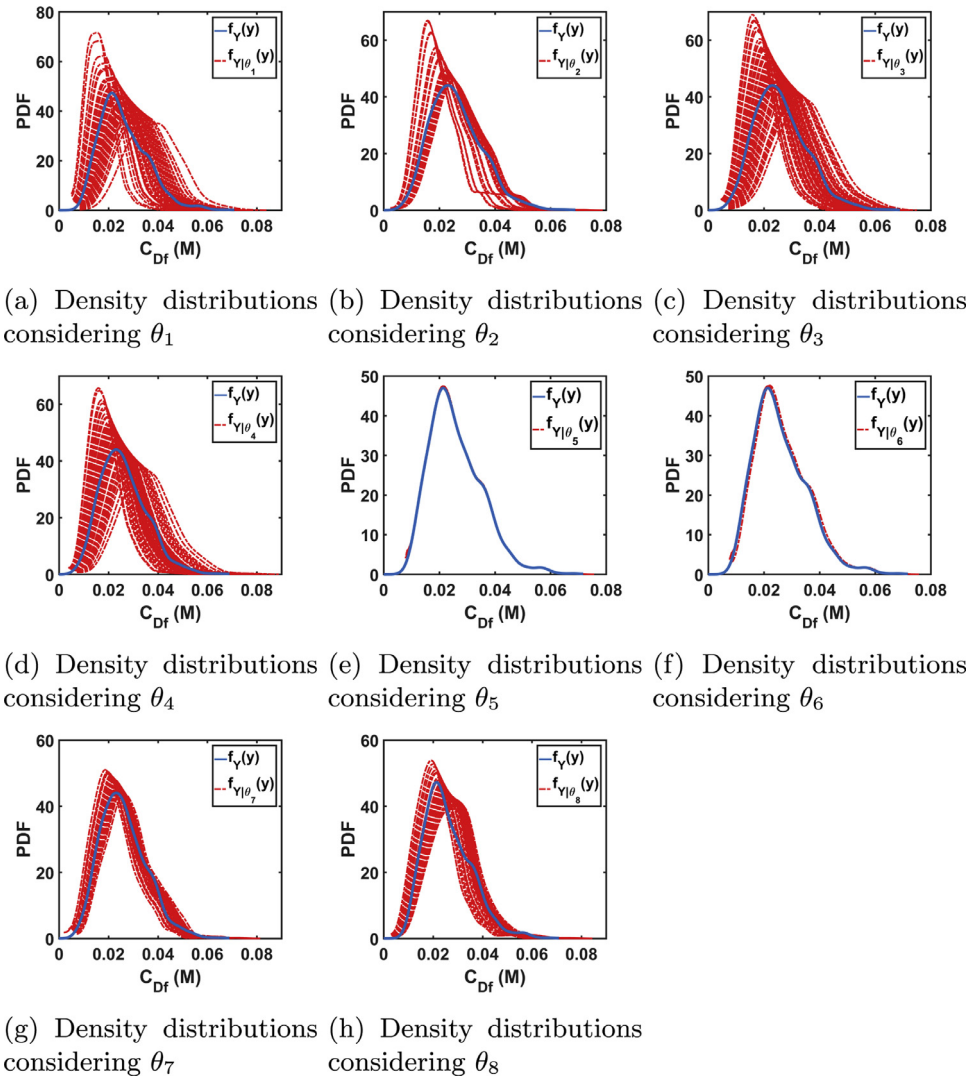
The SSI is obtained directly from the coefficients of the preceding PCE model by using Eqs. (48) and (49). The first-order and total sensitivities of the eight kinetic parameters for the final product concentration C_{Df} are summarized in Table 4. Evident differences exist among the magnitude of the sensitivities for different parameters. The first four parameters, θ_1, ...,θ_4, have the strongest impact on the final product concentration and its variance Var(C_{Df}), while the parameters θ_5, ...,θ_8 are less relevant. Moreover, the small deviation between the first-order, S_i^{uc}, and total sensitivities, $S_{T_i}^{uc}$, indicate that the interaction among the parameters is low, i.e, the sum of first-order sensitivities is close to 1.

Unlike the SSI calculated directly from the coefficients of the PCE model, MISA is computed by the method described in Section 4.4. The samples in the outer loop (n_o) and the inner loop (n_i) are set to 1000. Thus, a total number of 8×10^6 evaluations of the PCE model are required for calculating the indicators. Fig. 7 shows the comparison between the unconditioned and conditioned distributions of C_{Df}, where the effect of eliminating the uncertainty of one parameter can be directly observed in the corresponding sub-figures. The quantitative measures for independent parameters, i.e., indicator $\delta_i^{uc}(i = 1...8)$, are illustrated in Fig. 8a.

The sensitivity measures from the SSI and MISA reveal the influence of parameter uncertainties on the variation of C_{Df}. According to the results, we observe that the trends of the measures from the SSI and MISA are analogous. The kinetic parameters of reactions 1 and 2, i.e., θ_1, θ_2, θ_3 and θ_4, have higher influence than the others. This makes sense from a physical point of view as reactions 1 and 2 are faster than the other reactions and have a direct or indirect relation with product D. We can also observe that θ_1 and θ_3 have similar importance, which is also true for θ_2 and θ_4. The reason is that reactions 1 and 2 are parallel and have similar competitiveness. In contrast, the kinetic parameters of reactions 3 and 4, i.e.,θ_5, θ_6, θ_7 and θ_8, have lower influence, because

Table 4
First-order and total Sobol' sensitivity indices of the kinetic parameters for the final product concentration C_D.

Parameters	S_i^{uc}	$S_{T_i}^{uc}$
θ_1	0.2547	0.2684
θ_2	0.1790	0.1896
θ_3	0.2535	0.2673
θ_4	0.1786	0.1893
θ_5	2.3011×10^{-5}	4.3189×10^{-5}
θ_6	3.7996×10^{-5}	7.7143×10^{-5}
θ_7	0.0400	0.0514
θ_8	0.0577	0.0710
Sum	0.9637	–
Var(C_{Df})	8.5989×10^{-5}	



(a) Density distributions considering θ_1 (b) Density distributions considering θ_2 (c) Density distributions considering θ_3

(d) Density distributions considering θ_4 (e) Density distributions considering θ_5 (f) Density distributions considering θ_6

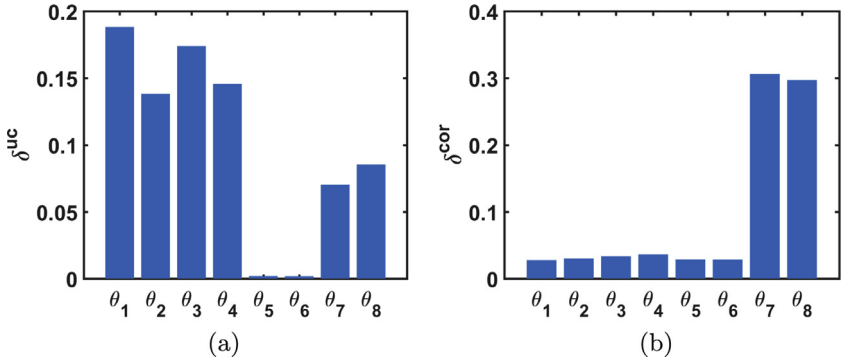
(g) Density distributions considering θ_7 (h) Density distributions considering θ_8

Fig. 7. Comparison of unconditioned distributions (blue line) and conditioned distributions (red lines) of concentrations of component D representing different parameter realizations in the absence of correlations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

reactions 3 and 4 are slow. However, θ_7 and θ_8 are more important than θ_5 and θ_6 because they have a direct impact on product D. As we can see, results from both methods are consistent with the structure of the model when the correlations among the parameters are neglected. In the following, the effect of the correlations on the results of the sensitivity analysis is presented.

5.4. Sensitivity measures in the presence of correlations

We investigate the effect of parameter correlations given in Table 2 on the results of the sensitivity analysis. According to the correlation matrix, the parameters are divided into three subgroups: (1) θ_1, θ_2, θ_3 and θ_4, (2) θ_5 and θ_6, and (3) θ_7 and θ_8. The parameters from the same



(a) (b)

Fig. 8. MISA sensitivity values of moment-independent measure δ^{uc} for the eight kinetic parameters when ignoring parameter correlations (a), and when parameter correlations are considered properly (b).

Table 5
Covariance-based sensitivity indices estimated with sparse PCE.

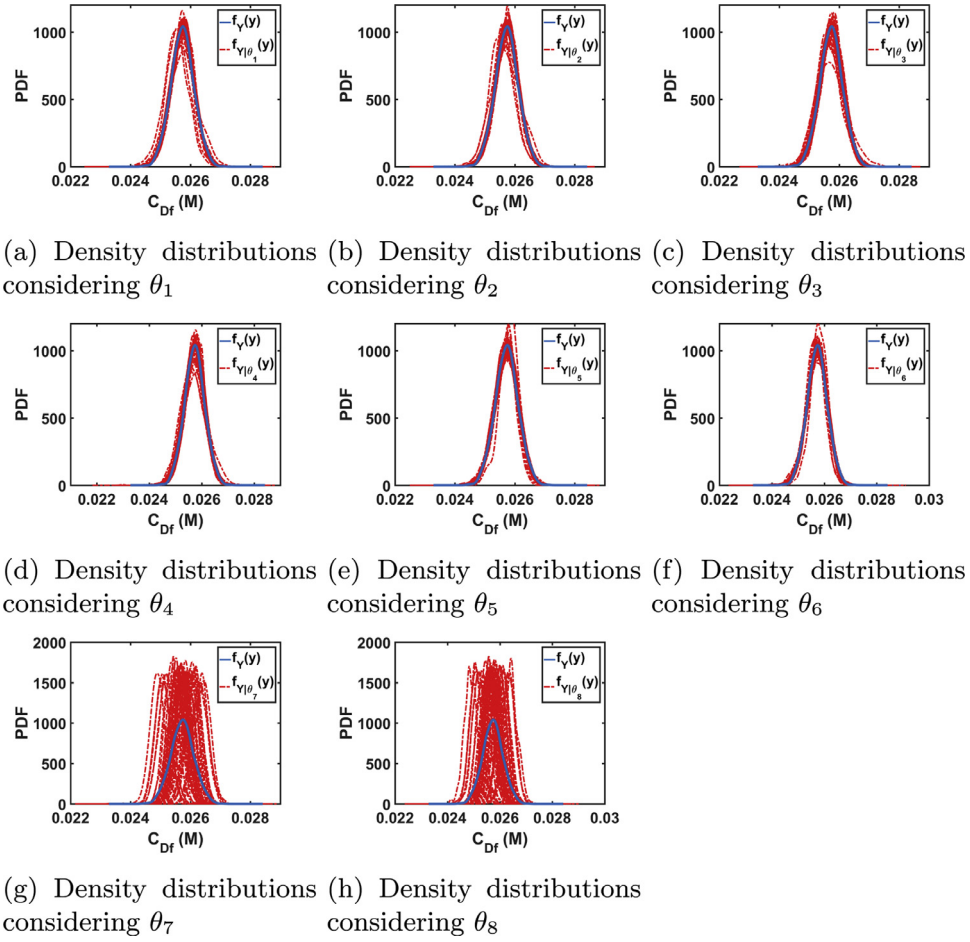
Parameters	S_i^U	S_i^C	S_i^{cov}	S_{Ti}^U	S_{Ti}^C	S_{Ti}^{cov}
θ_1	131.80	−132.98	−1.18	136.23	−137.47	−1.24
θ_2	92.72	−91.55	1.17	99.35	−98.15	1.19
θ_3	131.91	−129.88	2.03	141.63	−139.52	2.11
θ_4	93.09	−94.76	−1.67	96.30	−97.96	−1.66
θ_5	0.01	−0.01	−0.00	0.03	−0.03	0.00
θ_6	0.02	−0.02	0.00	0.02	−0.02	0.00
θ_7	20.68	−24.16	−3.48	20.73	−24.23	−3.50
θ_8	29.87	−25.57	4.30	26.27	−22.26	4.01
Sum	–	–	1.17	–	–	–
Var(Y)			1.65×10^{-7}			

subgroup have a strong correlation, while the parameters from different subgroups have a weak or even no correlation. The samples used in the following calculations are generated with Gaussian copula as Eq. (3).

Due to the presence of correlations among the parameters, the SSI is not well defined and, therefore, cannot be used in this situation. In contrast, CoDSA is still available to observe the effect of the parameters, as well as their correlations. The method presented in Section 4.3.2 is used to compute CoDSA with a sample size of 10,000. The calculated sensitivity indices are listed in Table 5, where S_i^U, S_i^C and S_i^{cov} are the first-order sensitivity indices and S_{Ti}^U, S_{Ti}^C and S_{Ti}^{cov} are the total sensitivity indices for the corresponding parameters, respectively. When comparing the first-order and total sensitivity indices, we observe that the interaction term is not relevant to describe the parameter influence on the model output. Moreover, S_i^U has the same trend as the SSI result but with different magnitudes. The large magnitudes of S_i^U mean the parameter uncertainties have a stronger influence on the model output if they are independent. This can also be observed from S_i^C which represents the effect of correlations and is negative in this case. It turns out the correlations reduce the importance of the parameters and the variance of the model output. The total impact of the parameters on the model output is indicated by the total covariance-based sensitivity indices S_i^{cov}. Please note that the existence of negative values for S_i^{cov} is due to the covariance function formulation and the importance of the parameter is quantified by the absolute value of S_i^{cov}.

In contrast to the variance-based SA methods, MISA is well posed in the presence of correlations among parameters, as its formulation is not based on the assumption of independent parameters. A similar structure for computation as for the independent case is used. Here, however, the correlation matrix of the joint density distribution of the parameters is added. In Fig. A.12a, we show the indicators δ_i^{cor} which are obtained for the given parameter correlations. Here, the most sensible parameters for C_{Df} are the kinetic parameters θ_7 and θ_8 for reaction 4, which is different from the case with independent parameters. A detailed analysis of the related probability distributions, see Fig. 9, explains the new parameter ranking. Here, in Fig. 9g and 9 h, the probability distribution of C_{Df} shifts dramatically if θ_7 and θ_8 are given. θ_1, θ_2, θ_3 and θ_4 still have a non-ignorable impact on the model response but are impaired by the correlation between the parameters. θ_5 and θ_6 have the weakest impact.

In this case study, we can conclude that the outcome of the parameter importance ranking is severely affected by the sensitivity measure we use; i.e., when applying the (co)variance-based or the moment-independent approach the parameter sensitivities are qualitatively similar. In contrast, the consideration of parameter correlations is crucial



(a) Density distributions considering θ_1 (b) Density distributions considering θ_2 (c) Density distributions considering θ_3

(d) Density distributions considering θ_4 (e) Density distributions considering θ_5 (f) Density distributions considering θ_6

(g) Density distributions considering θ_7 (h) Density distributions considering θ_8

Fig. 9. Comparison of unconditioned distributions (blue line) and conditioned distributions (red lines) of concentrations of component D considering different parameters in the presence of correlations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

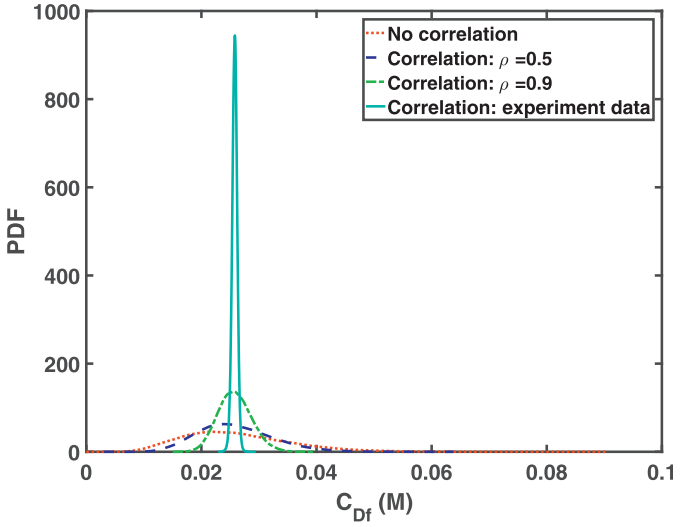


Fig. 10. Comparison of the resulting probability density functions of C_{Df} in the absence and presence of parameter correlations. The correlation coefficients for θ_1, θ_2, θ_3, θ_4, θ_7, θ_8 are assigned with 0.5 and 0.9 assuming factitious, equal correlations for two cases. The result based on a correlation matrix derived from experimental data is illustrated in addition.

Table 6
Parameter ranking from (co)variance-based and moment-independent sensitivity analysis in the presence and absence of correlations.

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
S_i^{uc}	1	3	2	4	8	7	6	5
S_i^{cor}	5	6	3	4	8	7	2	1
δ_i^{uc}	1	4	2	3	8	7	6	5
δ_i^{cor}	6	5	4	3	8	7	1	2

Table 7
Savage score correlation coefficients on importance ranking from variance-based and moment-independent sensitivity in the absence and presence of correlations.

SSCC	S^{uc}	S^{cov}	δ^{uc}	δ^{cor}
S^{uc}	1			
S^{cov}	−0.06	1		
δ^{uc}	0.98	−0.03	1	
δ^{cor}	−0.21	0.78	−0.17	1

for the parameter importance ranking. Thus, the impact of parameter correlations are further discussed in the next section.

5.5. Comparison of the results in the absence and presence of correlations

To compare the resulting variation of C_{Df} in the absence and presence of correlations, the corresponding probability distributions are illustrated in Fig. 10. To further demonstrate the effect of parameter correlation which are different from one and less dominating, two additional scenarios are shown in Fig. 10. As expected, the correlation has a considerable impact on the resulting distributions. Here, the spread of the shown distribution increases for lower correlation values but is still different compared to the nominal case, i.e., assuming no correlation. It turns out that the uncertainty of the model output, C_{Df}, which is estimated under the assumption of independent parameters for this study is magnified and, therefore, may render a model-based robust design strategy too conservative.

In this section, we compare the importance ranking for the different sensitivity measures neglecting the parameter correlation, as well as including parameter correlations based on experimental data; see Table 6. For the sake of completeness, results for fictitious correlation coefficients can be found in the Appendix A. The ranking from the first (third) and second (fourth) rows are transposed, while the rankings from first (second) and third (fourth) rows are analogous. To get a more quantitative comparison of the ranking, we compute the Savage score correlation coefficient (SSCC, [19]) for comparing the four sensitivity measures, see Table 7. The SSCC has a value range from -1 to 1, where 1 and -1 indicate identical and transposed rankings, respectively. As we can see from Table 7, the values of $SSCC(S^{uc}, \delta^{uc})$ and $SSCC(S^{cov}, \delta^{cor})$ are high and close to 1. However, the values of $SSCC(S^{uc}, S^{cov})$ and $SSCC(\delta^{uc}, \delta^{cor})$ are low and negative as the most relevant variables for them are different. According to this, we see that the discrepancy in the ranking of the most relevant parameters emerges due to the existence of parameter correlations and is less affected by the particular method used for the global sensitivity analysis.

6. Conclusion

In this paper, we presented different methods for global sensitivity analysis in the absence and presence of parameter correlations and compared them critically for the continuous synthesis of an active pharmaceutical ingredient scaffold. Sparse polynomial chaos expansion (PCE) was introduced for calculating these sensitivity measures efficiently. Gaussian copulas were utilized to sample from joint and conditional distributions, representing independent and correlated model parameters. In the case study, a continuous-flow reactor model was implemented and analyzed. PCE surrogate model was generated for this reactor model. Here, we saw that the presented least angle regression (LAR) algorithm improves the efficiency in PCE modeling. We also observed that the PCE model can approximate relevant statistics of simulation results at low computational cost. After performing the sensitivity analysis on the PCE model, the obtained results were compared between (co)variance-based methods (SSI and CoDSA) and the moment-independent method (MISA), assuming uncorrelated parameters. We observed a similar parameter sensitivity ranking for the analyzed metrics; that is, the parameters with a high impact on the model variation influence the output variance considerably, too. Moreover, we compared the results for independent and correlated parameters and saw that the parameter sensitivity ranking was quite different. For independent parameters, the kinetics of reactions 1 and 2 influenced the simulation outcome the most. For the correlated parameters, however, their impact was reduced significantly, and the kinetics of reaction 4 dominated. Moreover, we noticed that the variance and the width of the distribution of the model output were decreased once parameter correlations were considered. In summary, global sensitivities provide useful information for analysis and design in the field of chemical engineering and can be derived at acceptable computational cost even for complex problems when using PCE in combination with the LAR algorithm. MISA might be preferable because it is available for problems with independent and correlated model parameters. In addition, MISA is also more rigorous and precise than the (co)variance-based method as MISA considers the entire distribution instead of a single moment of the model output. However, independent of the metrics used for global parameter sensitivity, parameter correlations should always be considered, i.e., utilizing the full information of the parameter (co)variance matrix for process analysis and design. Further research for robust design of chemical processes under uncertainty reflecting parameter correlation will be conducted in the future.

Acknowledgment

We gratefully acknowledge the funding of the ”Promotionsprogramm μ-Props” by MWK Niedersachsen. The authors wish to thank Moritz Schulze and Victor Emenike for their helpful comments on the initial manuscript. Moreover, the anonymous reviewers deserve special thanks for their valuable input.

Appendix A. Additional results for sensitivity analysis with different correlation coefficients

The results of additional four cases with fictitious correlation coefficients are presented here, where the case 1,2,3, and 4 listed in Table A.8 are the same with the cases depicted in Figs. 10 andA.11.

As we mentioned in Section 5.5, the output variances depend considerably on the parameter correlation values. Although our primary focus was on correlation coefficients derived with experimental data, it might be still interesting to demonstrate the effect of parameter correlations which are less dominating; i.e., not that close to one. Here, we consider two additional test cases with fictitious correlation values: 0.5 and 0.9. CoDSA and MISA results are listed in Table A.8. Obviously, for both sensitivity measures the importance of parameters θ_7 and θ_8 increases gradually with higher correlation coefficients. The overall effect of θ_1, θ_2, θ_3, and θ_4 still dominates the output variation but recedes if the correlation coefficients increase further as the case with experimentally derived correlations. Furthermore, we include two additional cases 3 and 4, where the correlation coefficients for θ_1, θ_2, θ_3, θ_4 and θ_7, θ_8 are allocated with different values. The shapes of the output distribution are similar, see Fig. A.11. The sensitivity results,

Table A1
Sensitivity results for cases with different correlation coefficients. The correlation coefficients for θ_1, θ_2, θ_3, θ_4, θ_7, θ_8 are assigned with 0.5 and 0.9 for case 1 and 2, respectively. The correlation coefficients for θ_1, θ_2, θ_3, θ_4 and θ_7, θ_8 are assigned with 0.5 and 0.9 for the case 3, and with 0.9 and 0.2 for the case 4.

		θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
S^{cov}	1	0.258	0.184	0.260	0.182	0	0	0.030	0.062
	2	0.242	0.201	0.280	0.168	0	0	−0.026	0.124
	3	0.276	0.197	0.278	0.195	0	0	−0.005	0.026
	4	0.143	0.121	0.168	0.099	0	0	0.175	0.279
δ^{cor}	1	0.126	0.105	0.127	0.104	0.015	0.015	0.036	0.061
	2	0.054	0.048	0.058	0.045	0.016	0.014	0.021	0.057
	3	0.132	0.110	0.133	0.108	0.015	0.014	0.016	0.029
	4	0.045	0.042	0.047	0.038	0.016	0.017	0.117	0.161

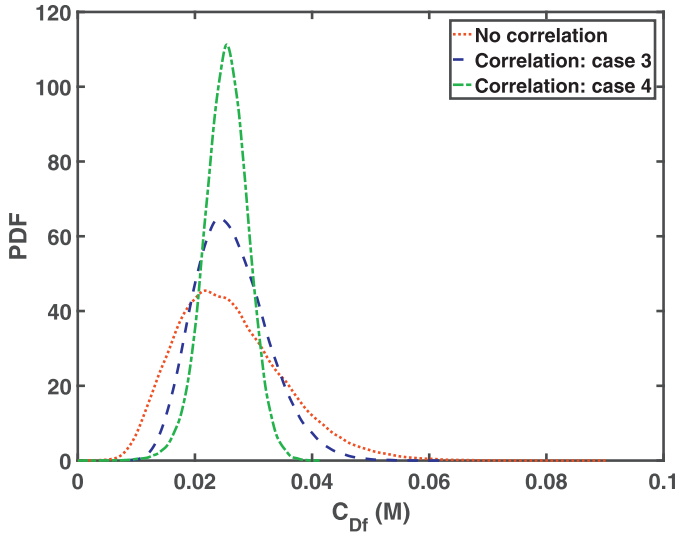


Fig. A1. Comparison of the resulting probability density functions of C_{Df} in the absence and presence of parameter correlations. The correlation coefficients for θ_1, θ_2, θ_3, θ_4 and θ_7, θ_8 are assigned with 0.5 and 0.9 for case 4, and with 0.9 and 0.2 for case 5.

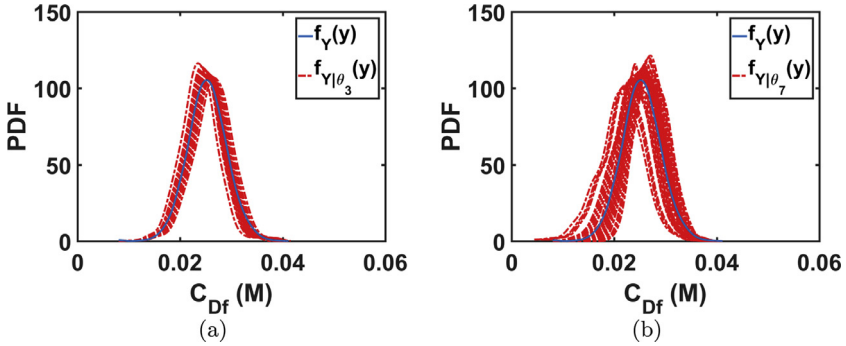


Fig. A2. Comparison of unconditioned distributions (blue line) and conditioned distributions (red lines) of concentrations of component D considering parameters θ_3 and θ_7 in case 5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

however, are completely different, especially for the most significant parameter. The sensitivity measure from CoDSA and MISA are analogue with minor differences. For instance, the sensitivity of θ_3 and θ_7 for case 5 are similar for CoDSA, while the sensitivity of θ_7 is twice as large compared to θ_3 for MISA. The reason is that MISA takes into account not only the output variance but also higher statistical moments, i.e., the entire output distribution. In Fig. A.12, we see that the shift of the conditional distributions is mainly due to the change of other moments, e.g., kurtosis, but not the variance which supports our conclusions considering MISA as a valuable tool in sensitivity analysis.

References

[1] Atherton R, Schainker R, Ducot E. On the statistical sensitivity analysis of models for chemical kinetics. AlChE J 1975;21(3):441–8.
[2] Biegler LT. Nonlinear programming: concepts, algorithms, and applications to chemical processes. SIAM; 2010.
[3] Blatman G. Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis. Clermont-Ferrand 2; 2009. Ph.D. thesis.
[4] Borgonovo E. A new uncertainty importance measure. Reliab Eng Syst Saf 2007;92(6):771–84.
[5] Borgonovo E, Apostolakis GE. A new importance measure for risk-informed decision making. Reliab Eng Syst Saf 2001;72(2):193–212.
[6] Borgonovo E, Plischke E. Sensitivity analysis: a review of recent advances. Eur J Oper Res 2016;248(3):869–87.
[7] Botev ZI, Grotowski JF, Kroese DP, et al. Kernel density estimation via diffusion. Ann Stat 2010;38(5):2916–57.
[8] Boukouvala F, Niotis V, Ramachandran R, Muzzio FJ, Ierapetritou MG. An integrated approach for dynamic flowsheet modeling and sensitivity analysis of a continuous tablet manufacturing process. Comput Chem Eng 2012;42:30–47.
[9] Campolongo F, Cariboni J, Saltelli A. An effective screening design for sensitivity analysis of large models. Environ Model Softw 2007;22(10):1509–18.
[10] Chu Y, Hahn J. Necessary condition for applying experimental design criteria to global sensitivity analysis results. Comput Chem Eng 2013;48:280–92.
[11] Crestaux T, Le Maître O, Martinez J-M. Polynomial chaos expansion for sensitivity analysis. Reliab Eng Syst Saf 2009;94(7):1161–72.
[12] Degerman M, Westerberg K, Nilsson B. Determining critical process parameters and process robustness in preparative chromatography–a model-based approach. Chem Eng Technol 2009;32(6):903–11.
[13] Efron B, Hastie T, Johnstone I, Tibshirani R, et al. Least angle regression. Ann Stat 2004;32(2):407–99.
[14] Englezos P, Kalogerakis N. Applied parameter estimation for chemical engineers. CRC Press; 2000.
[15] Ghanem RG, Spanos PD. Stochastic finite elements: a spectral approach. Courier Corporation; 2003.
[16] Haaker M, Verheijen P. Local and global sensitivity analysis for a reactor design with parameter uncertainty. Chem Eng Res Des 2004;82(5):591–8.
[17] Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab Eng Syst Saf 2003;81(1):23–69.
[18] Houska B, Logist F, Van Impe J, Diehl M. Robust optimization of nonlinear dynamic systems with application to a jacketed tubular reactor. J Process Control 2012;22(6):1152–60.
[19] Iman RL, Conover W. A measure of top–down correlation. Technometrics 1987;29(3):351–7.
[20] Lakerveld R, Benyahia B, Braatz RD, Barton PI. Model-based design of a plant-wide control strategy for a continuous pharmaceutical plant. AlChE J 2013;59(10):3671–85.
[21] Lebrun R, Dutfoy A. A generalization of the Nataf transformation to distributions with elliptical copula. Probab Eng Mech 2009;24(2):172–8.
[22] Li G, Rabitz H. General formulation of HDMR component functions with independent and correlated variables. J Math Chem 2012;50(1):99–130.
[23] Li G, Rabitz H. Relationship between sensitivity indices defined by variance-and covariance-based methods. Reliab Eng Syst Saf 2017;167:136–57.
[24] Li G, Rabitz H, Yelvington PE, Oluwole OO, Bacon F, Kolb CE, et al. Global sensitivity analysis for systems with independent and/or correlated inputs. J Phys Chem A 2010;114(19):6022–32.
[25] Li G, Rosenthal C, Rabitz H. High dimensional model representations. J Phys Chem A 2001;105(33):7765–77.
[26] Lin X, Janak SL, Floudas CA. A new robust optimization approach for scheduling under uncertainty:: i. bounded uncertainty. Comput Chem Eng 2004;28(6):1069–85.
[27] López-Benito A, Bolado-Lavín R. A case study on global sensitivity analysis with dependent inputs: the natural gas transmission model. Reliab Eng Syst Saf 2017;165:11–21.
[28] Luyben WL. Process modeling, simulation and control for chemical engineers. McGraw-Hill Higher Education; 1989.
[29] Mara TA, Tarantola S. Variance-based sensitivity indices for models with dependent inputs. Reliab Eng Syst Saf 2012;107:115–21.
[30] Mara TA, Tarantola S, Annoni P. Non-parametric methods for global sensitivity analysis of model output with dependent inputs. Environ Model Softw 2015;72:173–83.
[31] Marelli S, Sudret B. UQLab: A framework for uncertainty quantification in MATLAB. ETH-Zürich; 2014.
[32] Navarro M., Witteveen J., Blom J.. Polynomial chaos expansion for general multivariate distributions with correlated variables. arXiv:14065483 2014.
[33] Nelsen RB. An introduction to copulas. Springer Science & Business Media; 2007.
[34] Oakley JE, O'Hagan A. Probabilistic sensitivity analysis of complex models: a Bayesian approach. J R Stat Soc Ser B (Stat Methodol) 2004;66(3):751–69.
[35] Oladyshkin S, Nowak W. Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion. Reliab Eng Syst Saf 2012;106:179–90.
[36] Reizman BJ, Jensen KF. An automated continuous-flow platform for the estimation of multistep reaction kinetics. Organic Process Res Dev 2012;16(11):1770–82.
[37] Rosenblatt M. Remarks on a multivariate transformation. Ann Math Stat 1952;23(3):470–2.
[38] Sahinidis NV. Optimization under uncertainty: state-of-the-art and opportunities. Comput Chem Eng 2004;28(6):971–83.
[39] Saltelli A., Aleksankina K., Becker W., Fennell P., Ferretti F., Holst N., et al. Why so many published sensitivity analyses are false. a systematic review of sensitivity analysis practices. arXiv:171111359 2017.
[40] Saltelli A, Chan K, Scott EM, et al. Sensitivity analysis. 1. Wiley New York; 2000.
[41] Saltelli A, Ratto M, Tarantola S, Campolongo F. Sensitivity analysis for chemical models. Chem Rev 2005;105(7):2811–28.
[42] Saltelli A, Sobol IM. About the use of rank transformation in sensitivity analysis of model output. Reliab Eng Syst Saf 1995;50(3):225–39.
[43] Saltelli A, Tarantola S, Chan K-S. A quantitative model-independent method for global sensitivity analysis of model output. Technometrics 1999;41(1):39–56.
[44] Schenkendorf R. A general framework for uncertainty propagation based on point estimate methods. Proceedings of the Second European Conference of the Prognostics and Health Management Society, PHME14. Nantes, France; 2014.
[45] Sklar M. Fonctions de répartition à n dimensions et leurs marges. Université Paris 8; 1959.
[46] Sobol' IM. On sensitivity estimation for nonlinear mathematical models. Matematicheskoe Modelirovanie 1990;2(1):112–8.
[47] Sobol IM. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. Math Comput Simul 2001;55(1):271–80.
[48] Sobol IM, Kucherenko S. Derivative based global sensitivity measures and their link with global sensitivity indices. Math Comput Simul 2009;79(10):3009–17.
[49] Srinivasan B, Bonvin D, Visser E, Palanki S. Dynamic optimization of batch processes: ii. role of measurements in handling uncertainty. Comput Chem Eng 2003;27(1):27–44.
[50] Sudret B., Caniou Y.. Analysis of covariance (ANCOVA) using polynomial chaos expansions. Proceedings of the 11th International Conference on Structural Safety & Reliability2013;.
[51] Timan AF. Theory of approximation of functions of a real variable. 34. Elsevier; 2014.
[52] Turányi T. Sensitivity analysis of complex kinetic systems. tools and applications. J Math Chem 1990;5(3):203–48.
[53] Valkó É, Varga T, Tomlin A, Busai Á, Turányi T. Investigation of the effect of correlated uncertain rate parameters via the calculation of global and local sensitivity indices. J Math Chem 2017:1–26.
[54] Valkó É, Varga T, Tomlin A, Turányi T. Investigation of the effect of correlated uncertain rate parameters on a model of hydrogen combustion using a generalized HDMR method. Proc Combust Inst 2017;36(1):681–9.
[55] Vallerio M, Telen D, Cabianca L, Manenti F, Van Impe J, Logist F. Robust multi-objective dynamic optimization of chemical processes using the sigma point method. Chem Eng Sci 2016;140:201–16.
[56] Wan X, Karniadakis GE. Beyond Wiener–Askey expansions: handling arbitrary pdfs. J Sci Comput 2006;27(1–3):455–64.
[57] Wei P, Lu Z, Yuan X. Monte carlo simulation for moment-independent sensitivity analysis. Reliab Eng Syst Saf 2013;110:60–7.
[58] Wiener N. The homogeneous chaos. Am J Math 1938;60(4):897–936.
[59] Witteveen JA, Sarkar S, Bijl H. Modeling physical uncertainties in dynamic stall induced fluid–structure interaction of turbine blades using arbitrary polynomial chaos. Comput Struct 2007;85(11):866–78.
[60] Xiu D. Numerical methods for stochastic computations: a spectral method approach. Princeton University Press; 2010.
[61] Xiu D, Karniadakis GE. The wiener–askey polynomial chaos for stochastic differential equations. SIAM J Sci Comput 2002;24(2):619–44.
[62] Xu C, Gertner GZ. Uncertainty and sensitivity analysis for models with correlated parameters. Reliab Eng Syst Saf 2008;93(10):1563–73.
[63] Yang X, Mannan MS. The development and application of dynamic operational risk assessment in oil/gas and chemical process industry. Reliab Eng Syst Saf 2010;95(7):806–15.
[64] Zádor J, Zsely IG, Turányi T. Local and global uncertainty analysis of complex chemical kinetic systems. Reliab Eng Syst Saf 2006;91(10):1232–40.