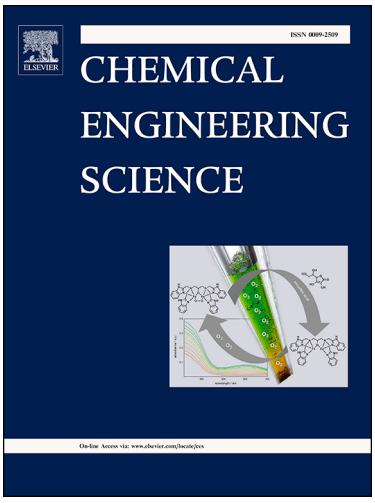


Robust optimization of a pharmaceutical freeze-drying process under non-Gaussian parameter uncertainties

Xiangzhong Xie, René Schenkendorf

PII: S0009-2509(19)30527-5
DOI: https://doi.org/10.1016/j.ces.2019.06.023
Reference: CES 15044

To appear in: *Chemical Engineering Science*

Received Date: 16 August 2018
Revised Date: 10 April 2019
Accepted Date: 17 June 2019

Please cite this article as: X. Xie, R. Schenkendorf, Robust optimization of a pharmaceutical freeze-drying process under non-Gaussian parameter uncertainties, *Chemical Engineering Science* (2019), doi: https://doi.org/10.1016/j.ces.2019.06.023

Robust optimization of a pharmaceutical freeze-drying process under non-Gaussian parameter uncertainties

Xiangzhong Xie[a,b], René Schenkendorf[a,b,*]

[a] *Technische Universität Braunschweig, Institute of Energy and Process Systems Engineering, Franz-Liszt-Straße 35, 38106 Braunschweig, Germany*
[b] *Technische Universität Braunschweig, Center of Pharmaceutical Engineering (PVZ), Franz-Liszt-Straße 35a, 38106 Braunschweig, Germany*

Abstract

Model-based design of pharmaceutical manufacturing processes has received much interest in academia and industry. Model parameter uncertainties, however, might deteriorate the predicted process performance. Probability-based robust process design concepts as a countermeasure against uncertainties might be implemented. Here, parameter uncertainties are typically limited to Gaussian parameter distributions. However, parameter uncertainties derived with experimental data can be correlated and arbitrarily distributed. In our previous work, transformation techniques were combined with the point estimate method (PEM) to address non-Gaussian and correlated parameter distributions, but at the cost of additional nonlinearities and approximation errors. In this work, we take advantage of Gaussian mixture distributions (GMD) and decompose the parameter distribution into a finite set of Gaussian distributions using the Expectation-Maximization approach. Combining the GMD with the PEM ensures a proper and effective uncertainty quantification. The improved PEM algorithm is applied to a freeze-drying process (lyophilization) aiming for high-quality products with minimum processing time. Results obtained suggest that the novel GMD-PEM algorithm has the potential to outperform conventional robustification concepts regarding credibility and efficiency.

*Corresponding author
Email address: r.schenkendorf@tu-braunschweig.de (René Schenkendorf)

Keywords: robust optimization, freeze-drying, point estimate method, Gaussian mixture distribution, uncertainty quantification, expectation-maximization

1. Introduction

The global competition in the (bio)chemical and pharmaceutical industries necessitates efficient and reliable manufacturing processes. Mathematical models, which mimic manufacturing processes, provide options for predicting and for optimizing the efficiency and effectiveness of the process performance with the desired reliability levels and low investment costs (Liu & Xu, 2017; Edgar et al., 2001; Biegler, 2010). Mathematical models for (bio)chemical and pharmaceutical processes typically consist of conservation equations and model parameters that describe the respective physiochemical phenomena (Luyben, 1989; Roffel & Betlem, 2007; Benyahia et al., 2012). However, due to model simplifications and measurement imperfections, the estimated model parameters are imprecise (Emery & Nenarokomov, 1998; Schenkendorf et al., 2018). Uncertain model parameters, in turn, lead to variations in model simulations that might result in performance losses and high-risk process operations (Grossmann & Sargent, 1978; Sahinidis, 2004; Telen et al., 2015; Vallerio et al., 2015). In the literature, the design of (bio)chemical processes considering model parameter uncertainties is known as robust process design.

1.1. Review of probability-based robust process design concepts

Robust process design intends to improve the robustness of the processes under model uncertainties at the cost of reduced process performance (Vallerio et al., 2015). For instance, scenario- and possibility-based approaches might be implemented to handle the adverse effect of parameter uncertainties making use of worst-case scenarios and fuzzy-set theory, respectively (Beyer & Sendhoff, 2007). Alternatively, probability-based approaches have been implemented extensively in various studies to solve robust optimization problems(Vallerio

et al., 2016; Shi et al., 2016; Sahinidis, 2004). These approaches provide an explicit description of uncertainties in terms of probability distributions, and thus, probability-based approaches ensure less conservative solutions compared with scenario- and possibility-based concepts (Sahinidis, 2004; Beyer & Sendhoff, 2007; Mesbah et al., 2014). However, probability-based approaches suffer from high computational costs of the propagation and quantification of uncertainties (Smith, 2013; Telen et al., 2015; Vallerio et al., 2016). A general probability-based robust optimization strategy, in which the point estimate method (PEM) was implemented and adapted further to correlated parameter distributions, are introduced and explained in our previous work (Schenkendorf, 2014a; Xie et al., 2018). The implemented PEM aims to avoid a vast amount of non-deterministic samples and associated simulation runs by analyzing exclusively statistical moments of low order. In doing so, fewer sample points compared with Monte Carlo simulations (MCs) and polynomial chaos expansion (PCE) have to be evaluated in robust process design (Oladyshkin & Nowak, 2012; Nimmegeers et al., 2016). Moreover, the PEM is more accurate than other cubature methods as concluded by Maußner & Freund (2018). The original PEM, however, is limited to Gaussian probability distributions (Lerner, 2002).

When utilizing experimental data for calibrating models, regression methods that assume Gaussian likelihood functions are typically used, and the inverse of the Fisher information matrix (FIM) is the standard for quantifying parameter uncertainties. This concept of uncertainty quantification, however, is accurate only if the process model is linear regarding the model parameters, and the parameter uncertainties are Gaussian distributed (Shi et al., 2014; Schenkendorf et al., 2018). Practically, the amount of experimental data of novel (bio)chemical and pharmaceutical products is always limited, and the applied process models are highly nonlinear and complex (Sheehan & Liapis, 1998; Carullo & Vallan, 2012). Shi et al. (2014) compared estimated parameter uncertainties based on the regression method and a Bayesian technique. It was found that the shapes of the probability distribution of the model parameters are generally non-Gaussian for nonlinear models. Kalyanaraman et al. (2015) also obtained non-Gaussian

distributions for parameter uncertainty estimated with experimental data using the Bayesian approach. A similar conclusion is given in Joshi et al. (2006) where the parameter uncertainties, which are approximated by the bootstrap method, have non-Gaussian distributions. Therefore, it is necessary to consider probability distributions with non-Gaussian distributions for robust process design in the field of (bio)chemical engineering (Rossner, 2014; Xu & Gómez-Hernández, 2016; Oladyshkin & Nowak, 2012). Robust process design that includes parameter uncertainties of non-Gaussian, data-centric probability density functions is still missing in the literature and is the focus of this work. To apply the PEM for these non-Gaussian uncertainties, we present an adapted sampling scheme within a Gaussian-mixture framework for proper uncertainty quantification.

Commonly, not only the non-Gaussian shape of parameter distributions but also parameter correlations are ignored to simplify robust process design problems (Telen et al., 2015; Maußner & Freund, 2018). This simplifying assumption, in turn, leads to an inevitable loss of information and might result in sub-optimal process designs. Alternatively, in the case of the PEM, non-Gaussian distributions including parameter correlations can be incorporated via a transformation step; i.e., sample points from the original PEM are mapped with a single transformation function to mimic the correlated non-Gaussian distribution. In our previous work, we used the Gaussian copula and the Nataf transformation to map the original PEM samples to non-Gaussian distributions of interest, so that the information included in the non-Gaussian distribution, as well as the parameter correlation, is appropriately considered Xie et al. (2018). The transformation step, however, leads to extra complexity in the approximation which might reduce the accuracy of the PEM. Moreover, the Gaussian copula is capable only of describing a linear correlation between the parameters. In this work, we propose to use Gaussian mixture distributions (GMD) to represent non-Gaussian and correlated parameter uncertainties, i.e., the weighted sum of a limited number of Gaussian distributions. The GMD concept does not add any extra complexity in the approximation and is available for nonlinear parameter correlation problems.

GMD has been extensively applied in the field of pattern recognition and machine learning to cluster data into subgroups (Bishop, 2006). Rossner (2014) came up with the idea of using GMD to decompose a one-dimensional Gaussian distribution into several component distributions to approximate the uncertainties in the model output more accurately even if the resulting model uncertainties are non-Gaussian. Technically, a least square estimation approach with additional constraints on the width of the Gaussian mixture components was used to determine the weight factors of the GMD. In this work, we also use GMD for proper uncertainty quantification but follow a different philosophy. First, we decompose the multivariate distribution directly instead of doing that individually for each marginal distribution which leads to lower computational costs. Second, we implement the expectation-maximization (EM) algorithm to cluster the data and to estimate the weight factors of the GMD. Note that for the traditional least square estimation as suggested in Rossner (2014), the optimal weight factors are difficult to obtain, especially for high-dimensional parameter problems (Ng et al., 2012). With this novel GMD-PEM algorithm for robust process design, we can capture the shapes of the non-Gaussian parameter distributions and the respective model output distributions realistically.

1.2. Background of the case study: freeze-drying process

The novel GMD-PEM algorithm in robust process design is motivated and applied to the freeze-drying process. The freeze-drying process, also named lyophilization, plays an essential role in (bio)pharmaceutical manufacturing to stabilize active pharmaceutical ingredients (APIs) which are unstable and have limited storage time in aqueous solutions, e.g., therapeutic protein formulations and vaccines (Amorij et al., 2007; Kasper et al., 2013). To minimize degradation effects and to ensure cake integrity, freeze-drying, which is a high-cost, time-consuming batch-wise process, requires careful formulation and process design. Here, primary drying where water is removed from the frozen product is the most time-consuming and failure-prone step (Fissore et al., 2010). As APIs used for freezing-drying are highly valuable products, the model-based design of

the primary drying step may help to ensure high quality at a competitive cost. A mathematical model for primary drying was investigated in Fissore et al. (2010, 2011) to mimic the mass and energy transfers during the freeze-drying process. Mortier et al. (2016) adapted the model for first steps in model-based process design. A grid-based approach was implemented to increase the efficiency of the primary drying step while guaranteeing critical quality attributes (CQA) of the dried product ICH (2009). Moreover, Monte Carlo simulations were also used to quantify the effect of parameter uncertainties which might result from measurement noise and model simplifications. However, optimization with the Monte Carlo simulations at individual time points might lead to a suboptimal solution and is computationally expansive by definition, especially in combination with Monte Carlo simulations for uncertainty quantification.

1.3. Scope of the present work

In this work, we aim to optimize the freeze-drying process under uncertain model parameters. Here, uncertainties with non-Gaussian distributions are assumed for two critical process parameters: the mass transfer resistance coefficient of the product and the heat transfer coefficient of the vial and the product (Tang & Pikal, 2004; Mortier et al., 2016). The non-Gaussian distributed parameters are taken into account in the freeze-drying simulation, and the resulting uncertainties in the sublimation mass and the temperature at the sublimation interface are quantified. For instance, the uncertain model parameters and the resulting variations in the sublimation mass are depicted in Fig. 1. The novel GMD-PEM algorithm is implemented to efficiently describe and quantify the non-Gaussian uncertainties in the model parameters and outputs, and these uncertainties are integrated into the robust process design to ensure product quality standards and process efficiency simultaneously. Moreover, we demonstrate the superiority of the novel GMD-PEM algorithm by comparing the derived robust process design results with the original PEM and the PEM with the nonlinear transformation step.

The remainder of the paper is organized as follows. The mathematical back-

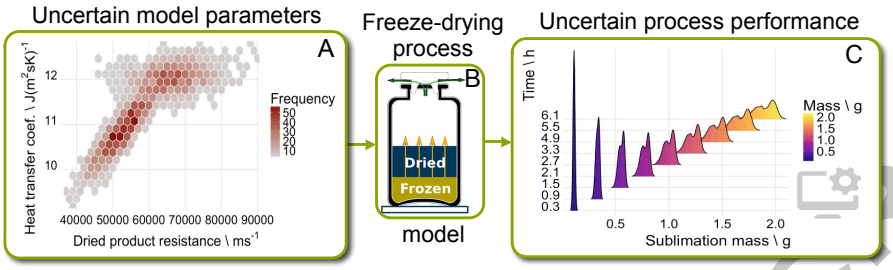


Figure 1: Illustration of uncertain model parameters in the freeze-drying process: Two correlated model parameters with a non-Gaussian distribution, see the scatter plot **(A)**, of the freeze-drying process **(B)** lead to uncertain non-Gaussian simulation results, e.g., sublimation mass distributions **(C)**.

ground of the methods and the structure of the robust process design problem are introduced in Section 2. The first-principle model of the freeze-drying process, as well as the assumed parameter uncertainties, is presented in Section 3. Results and discussion about the performance of the proposed GMD-PEM algorithm for robust process design are provided in Section 4. In Section 5, the conclusions are given.

2. Methodology

In this section, we describe the basics and the mathematical formulations of the GMD-PEM algorithm for robust process design. First, we start with a brief review of the PEM, the single Gaussian approach, and the transformed arbitrary distribution approach. Second, we introduce the EM algorithm to calibrate the GMD-PEM algorithm, i.e., to iteratively determine the maximum likelihood of the parameters even when the actual structure of the Gaussian mixture distribution is unknown.

Robust process design requires statistical information regarding the quantities of interests, e.g., yield, conversion and costs, to ensure meaningful designs based on simulation studies. To this end, the parameter uncertainties that are propagated through the model have to be transferred to the simulation results,

and the resulting statistics are then quantified. Two options are available to calculate the needed statistical information. One option is to estimate the probability distribution directly with Monte Carlo simulations and kernel density estimators (Mooney, 1997), which requires a vast number of samples to cover the relevant parameter space (Botev et al., 2010). The second option is to calculate statistical moments instead, e.g., mean, variance, skewness, and kurtosis, and to parameterize the probability distributions with these statistical moments (Zhao & Ono, 2001). As the statistical moments can be approximated with the PEM (Lerner, 2002; Schenkendorf, 2014a; Xie et al., 2018), the computational cost is typically lower compared to the direct approximation of the resulting probability density functions. However, for the same model, the accuracy of the approximated statistical moments decreases with the increase of the order of the statistical moment. Moreover, recovering the complete information of a non-Gaussian probability distribution with low-order statistical moments introduces additional approximation errors, which is further illustrated subsequently in the case study. The basics of the PEM are summarized in what follows.

2.1. Point estimate method (PEM)

Assuming an n_ξ-dimensional random parameter vector $\boldsymbol{\xi} \in I_{\boldsymbol{\xi}} \subset \mathbb{R}^{n_\xi}$, the corresponding multivariate normal distribution with the joint density function $\boldsymbol{p}(\boldsymbol{\xi})$ reads as:

$$\boldsymbol{p}(\boldsymbol{\xi}) = \frac{1}{\sqrt{(2\pi)^{n_\xi}|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{\xi} - \mu_{\boldsymbol{\xi}})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\xi} - \mu_{\boldsymbol{\xi}})\right), \tag{1}$$

where $\mu_{\boldsymbol{\xi}}$ and $\boldsymbol{\Sigma}$ are the vector of the mean values and the covariance matrix, respectively. Moreover, we assume a nonlinear function $\boldsymbol{f}(\boldsymbol{\xi}) : \mathbb{R}^{(n_\xi)} \rightarrow \mathbb{R}^{n_x}$ which represents the mathematical process model. The process model maps the model parameters to the model outputs; i.e., $\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{\xi})$. The n-th statistical moment of $\boldsymbol{x}$ is given as:

$$\boldsymbol{\mu}_n = \int_{I_{\boldsymbol{\xi}}} \boldsymbol{x}\boldsymbol{p}(\boldsymbol{\xi}) d\boldsymbol{\xi} \qquad n = 1, \tag{2}$$

$$\boldsymbol{\mu}_n = \int_{I_{\boldsymbol{\xi}}} (\boldsymbol{x} - \boldsymbol{\mu}_1)^n \boldsymbol{p}(\boldsymbol{\xi}) d\boldsymbol{\xi} \qquad n \geq 2. \tag{3}$$

The PEM can be used to approximate the integral terms in Eqs. (2) and (3) with the weighted sum of the model simulations which are evaluated at deterministic parameter sample points (Schenkendorf, 2014a):

$$\boldsymbol{\mu}_1 \approx w_0 \boldsymbol{f}(GF[0]) + w_1 \sum \boldsymbol{f}(GF[\pm\vartheta]) + w_2 \sum \boldsymbol{f}(GF[\pm\vartheta, \pm\vartheta]), \tag{4}$$

$$\begin{aligned} \boldsymbol{\mu}_n \approx\ & w_0(\boldsymbol{f}(GF[0]) - \boldsymbol{\mu}_1)^n + w_1 \sum (\boldsymbol{f}(GF[\pm\vartheta]) - \boldsymbol{\mu}_1)^n + \\ & w_2 \sum (\boldsymbol{f}(GF[\pm\vartheta, \pm\vartheta]) - \boldsymbol{\mu}_1)^n, \end{aligned} \tag{5}$$

where $\vartheta = \sqrt{3}, w_0 = 1 + \frac{n_\xi^2 - 7n_\xi}{18}, w_1 = \frac{4-n_\xi}{18}, w_2 = \frac{1}{36}$. The generator function $GF(\cdot)$ is used to create deterministic sample points used in Eq. (4) and (5) (Lerner, 2002). For instance, with $GF[\pm\vartheta]$ two sample points are generated starting with the first element of the original parameter vector that is assigned a new value with $+\vartheta$ and $-\vartheta$, respectively. The remaining elements of the parameter vector are processed similarly according to this permutation scheme. Please note, that for $GF[\pm\vartheta, \pm\vartheta]$ two elements of the parameter vector are permuted at the same time. So the number of sample points generated with $GF(\cdot)$ depends on the number of possible permutations. Theoretically, the total number of sample points from $GF[0], GF[\pm\vartheta], GF[\pm\vartheta, \pm\vartheta]$ is $2n_\xi^2 + 1$; i.e., the problem scales quadratically with the number of uncertain model parameters (Lerner, 2002).

As pointed out by Lerner (2002), the PEM can calculate integral terms of monomials only up to order of 5 accurately. In other words, the complexity of function $\boldsymbol{f}(\boldsymbol{\xi})$, as well as the order of the calculated statistical moments, determines the accuracy of Eqs. (4) and (5) when the original PEM is used. According to the discussion in Schenkendorf (2014b) and Maußner & Freund (2018), satisfactory estimations of the first- and second-order moments can be found for many engineering problems, while higher-order moments with $n > 2$ might be beyond the capability of the PEM. Note that the approximation with the above given $GF(\cdot)$ configuration is available only if $\boldsymbol{\xi}$ follows a standard multivariate normal distribution; i.e., $\mu_{\boldsymbol{\xi}} = \mathbf{0}$ and $\boldsymbol{\Sigma} = I$ (Lerner, 2002; Xie et al., 2018). Thus, for non-Gaussian parameter uncertainties, alternative concepts

have to be applied.

2.2. Non-Gaussian parameter uncertainties

2.2.1. Methods from literature and previous work

As mentioned in Rossner (2014), the distribution of parameter uncertainties estimated with experimental data from various resources might have arbitrary shapes; i.e., the parameter uncertainties cannot be described properly by Gaussian density functions. Thus, Eqs. (4) and (5) are ill-posed and have to be refined. To employ the PEM for non-Gaussian parameter distributions, the deterministic sample points have to be modified as illustrated in Fig. 2. Here, we assume $\boldsymbol{\theta} \in I_{\boldsymbol{\theta}} \subset \mathbb{R}^{n_\theta}$ is the vector of the parameters with non-Gaussian uncertainties, and we want to estimate the mean and variance of $\boldsymbol{f}(\boldsymbol{\theta})$ which could be the cost function for the process design or a constraint function as explained in the following section. The first idea, which is what we call the single Gaussian approach in the left column, is to approximate the arbitrary distribution with a multivariate normal distribution of which the mean $\mathbb{E}(\boldsymbol{\theta})$ and variance $\Sigma(\boldsymbol{\theta})$ according to:

$$\mathbb{E}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\theta}_i, \tag{6}$$

$$\Sigma(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} (\boldsymbol{\theta}_i - \mathbb{E}(\boldsymbol{\theta}))(\boldsymbol{\theta}_i - \mathbb{E}(\boldsymbol{\theta}))^T. \tag{7}$$

Note that the off-diagonal elements $\Sigma(\boldsymbol{\theta})$ are typically neglected and set to zero; i.e., the parameter uncertainties are considered a multivariate Gaussian distribution without parameter correlations. The original PEM samples generated by function $GF[\cdot]$ are mapped onto the the multivariate normal distribution described by $\mathbb{E}(\boldsymbol{\theta})$ and $\Sigma(\boldsymbol{\theta})$ with the linear transformation given below:

$$\boldsymbol{\theta}_i^{PEM} = A\boldsymbol{\xi}_i^{PEM} + \mathbb{E}(\boldsymbol{\theta}), \qquad i = 1, \ldots, 2n_\theta^2 + 1, \tag{8}$$

where A is the lower triangle matrix from the Cholesky decomposition of $\Sigma(\boldsymbol{\theta})$. According to the proposition in Xie et al. (2018), the mean and variance of $\boldsymbol{f}(\boldsymbol{\theta})$

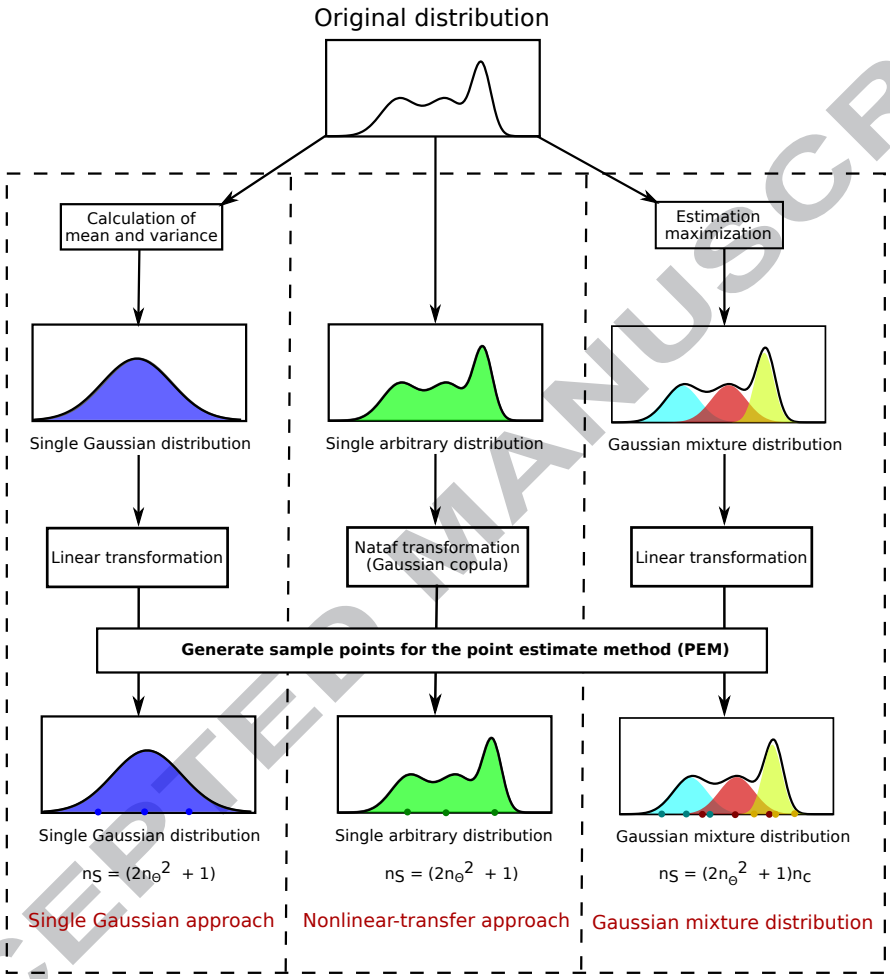


Figure 2: A schematic diagram for the single Gaussian approach, the nonlinear-transfer approach (previous work), and the Gaussian mixture distribution (GMD, present work) with an example of a one-dimensional probability distribution. n_S and n_θ are the number of sample points for the PEM and the number of parameters. n_c is the number of component mixtures for the GMD.

can be derived with the transformed samples based on Eqs. (4) and (5):

$$\boldsymbol{\mu}_1(\boldsymbol{f}(\boldsymbol{\theta})) \approx \sum_{i=1}^{2n_\theta^2+1} w_i \boldsymbol{f}(\theta_i^{PEM}), \tag{9}$$

$$\boldsymbol{\mu}_2(\boldsymbol{f}(\boldsymbol{\theta})) \approx \sum_{i=1}^{2n_\theta^2+1} w_i (\boldsymbol{f}(\theta_i^{PEM}) - \boldsymbol{\mu}_1(\boldsymbol{f}(\boldsymbol{\theta})))^2. \tag{10}$$

For the sake of simplicity, in the robust process design literature, it is commonly assumed that the parameter uncertainties follow a Gaussian distribution and are independent (Srinivasan et al., 2003; Telen et al., 2015; Shi et al., 2016; Kaiser et al., 2016). However, this assumption leads to a certain loss of information regarding parameter uncertainties and includes extra deviations in the approximations of Eqs. (9) and (10).

The second approach is one that we used in previous work Xie et al. (2018), the so-called nonlinear-transfer approach (NTA). It is illustrated in the middle column of Fig. 2. Instead of using a Gaussian distribution, we directly map the PEM samples to the actual parameter distribution. To this end, the iso-probabilistic transformation given in Eq. (11) is frequently used:

$$\boldsymbol{\theta}_i^{PEM} = F_{\boldsymbol{\theta}}^{-1}(F_\xi(\boldsymbol{\xi}_i^{PEM})). \tag{11}$$

In Eq. (11), $F_{\boldsymbol{\theta}}$ and $F_{\boldsymbol{\xi}}$ are the joint cumulative density function (CDF) for parameters $\boldsymbol{\theta}$ and standard Gaussian random parameters $\boldsymbol{\xi}$. The joint CDF $F_{\boldsymbol{\theta}}$ for multivariate distributions are usually complex and implicit, especially when the distributions are non-Gaussian and correlated. Moreover, it is also challenging to map directly from one parameter space to the other with Eq. (11) due to the complexity of $F_{\boldsymbol{\theta}}$. Therefore, $F_{\boldsymbol{\theta}}$ is approximated using Gaussian copulas with marginal CDFs F_{n_θ} for individual parameters and the correlation matrix Σ_ρ, as shown in Eq. (12) (Nelsen, 2007):

$$F_{\boldsymbol{\theta}} = F_N[F^{-1}(F_1), \cdots, F^{-1}(F_{n_\theta}); \Sigma_\rho]. \tag{12}$$

Here, F^{-1} denotes the inverse CDF of the standard normal distribution, F_N denotes the joint CDF of multivariate normal distributions, and $F_i, i = 1, \ldots, n_\theta$

are the marginal CDFs of parameters θ_i. In this study, the inverse Nataf transformation, which is an alternative numerical algorithm based on Eqs. (11) and (12), is used to transform the PEM samples of a standard multivariate Gaussian distribution to the samples of the target distribution of $\boldsymbol{\theta}$ (Noh et al., 2009; Lebrun & Dutfoy, 2009). With the modified samples, Eqs. (9) and (10) can be used to calculate the mean and the variance of the model outputs.

This method retains almost all of the parameter distribution information in the modified PEM sample points. However, the iso-probabilistic transformation might be highly nonlinear and adds extra complexity in the simulation results $\boldsymbol{f}(\boldsymbol{\theta}) = \boldsymbol{f}(F_{\boldsymbol{\theta}}^{-1}(F_{\xi}(\boldsymbol{\xi})))$. Note that the Gaussian copula is correct only when the parameters are linearly correlated. These two simplifying assumptions might result in sub-optimal robust process designs.

2.2.2. Gaussian mixture distributions (GMD)

GMD are an essential part of the proposed algorithm for robust process design. Unlike the two previous approaches where we either simplify a non-Gaussian distribution into an independent Gaussian distribution or transform the original PEM samples to a non-Gaussian distribution with nonlinear transformation functions, the GMD concept represents a non-Gaussian distribution with the superposition of a limited number of Gaussian distributions. The PEM samples are mapped to these Gaussian distributions with a linear transformation step, as illustrated in the right column in Fig. 2.

The GMD is structured as

$$\boldsymbol{\theta} \sim \sum_{j=1}^{N_C} \omega_j \mathcal{N}(\boldsymbol{\mu}_{1j}, \boldsymbol{\Sigma}_j), \tag{13}$$

where $\mathcal{N}(\boldsymbol{\mu}_{1j}, \boldsymbol{\Sigma}_j)$ means the Gaussian distribution with mean $\boldsymbol{\mu}_{1j}$ and covariance matrix $\boldsymbol{\Sigma}_j$ for the jth component. Here, ω_j is the non-negative weight for jth component with $\sum_{j=1}^{N_C} \omega_j = 1$. N_C is the total number of Gaussian distributions. The probability density function (PDF) of $\boldsymbol{\theta}$ is equal to:

$$\boldsymbol{p}(\boldsymbol{\theta}) = \sum_{j=1}^{N_C} \omega_j \boldsymbol{p}_j(\boldsymbol{\theta}), \tag{14}$$

where $\boldsymbol{p}_j(\boldsymbol{\theta})$ is the PDF of a Gaussian distribution given as

$$\boldsymbol{p}_j(\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^{n_\theta}|\boldsymbol{\Sigma}_j|}} exp(-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_{1j})^T \boldsymbol{\Sigma}_j{}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}_{1j})). \tag{15}$$

Note that the sum of random variables and the weighted sum of the Gaussian distributions are different concepts, i.e., the sum of the random variables results in a Gaussian distribution, whereas the weighted sum of the Gaussian distributions can represent non-Gaussian distributions needed for robust process design. With the approximation in Eq. (13), the original PEM samples are mapped individually to the component distributions with the linear transformation given in Eq. (8). Here, $N_C \times (2n_\theta^2 + 1)$ samples are obtained as:

$$\boldsymbol{\theta}_{ji}^{PEM} = A_j \boldsymbol{\xi}_i^{PEM} + \boldsymbol{\mu}_{1j}, \qquad i = 1 \ldots 2n_\theta^2 + 1, \quad j = 1 \ldots N_C, \tag{16}$$

where A_j is the lower triangle matrix from the Cholesky decomposition of the full co-variance matrix $\boldsymbol{\Sigma}_j$. To determine the samples with Eq. (16), we still need information about the structure of the GMD in Eq. (13). In other words, the values for ω_j, $\boldsymbol{\mu}_{1j}$, $\boldsymbol{\Sigma}_j$, and N_C have to be estimated.

Let us assume $\boldsymbol{y}$ are the realizations from a non-Gaussian distribution. Generally, the unknown parameters ω_j, $\boldsymbol{\mu}_{1j}$, and $\boldsymbol{\Sigma}_j$ can be determined by maximizing the marginal likelihood function of the unknown parameters with given realizations of $\boldsymbol{y}$ as:

$$\theta_p^* = \underset{\theta}{\operatorname{argmax}}\, L(\boldsymbol{y}; \theta_p), \tag{17}$$

$$L(\boldsymbol{y}; \theta_p) = \int L(\boldsymbol{y}, \boldsymbol{z}; \theta_p) d\boldsymbol{z}, \tag{18}$$

where θ_p is the vector of unknown parameters, i.e., ω_j, $\boldsymbol{\mu}_{1j}$, and $\boldsymbol{\Sigma}_j$, and L is the likelihood function (Kleinbaum & Klein, 2010; Rossner, 2014). The discretized variables $\boldsymbol{z}$ indicate the membership of each realization in one of the component distributions. Please note that $\boldsymbol{z}$ are unobserved variables and as such latent variables. Thus, solving the maximization problem in Eq. (18) is challenging (Bishop, 2006). Alternatively, the EM approach, which is commonly used to maximize the marginal likelihood function of parameters in cases where

latent variables exist (Bishop, 2006), can be used to estimate the parameters θ_p. Before implementing the EM approach, the likelihood function in Eq. (17) is reformulated as shown in Eq. (19). Here, the likelihood function $L(\boldsymbol{y};\theta_p)$ is written as a probability density function of realizations $\boldsymbol{y}$ conditioned on θ_p, i.e., $p(\boldsymbol{y}|\theta_p)$.

$$\begin{aligned}\ln(p(\boldsymbol{y}|\theta_p)) &= \int q(\boldsymbol{z}) \ln\left(\frac{p(\boldsymbol{y},\boldsymbol{z}|\theta_p)}{q(\boldsymbol{z})}\right) d\boldsymbol{z} - \int q(\boldsymbol{z}) \ln\left(\frac{p(\boldsymbol{z}|\boldsymbol{y};\theta_p)}{q(\boldsymbol{z})}\right) d\boldsymbol{z}, \\ &= F(q,\theta_p) + KL(q||p), \end{aligned} \tag{19}$$

where $KL(\cdot)$ is the non-negative Kullback-Leibler divergence measure. And $F(q,\theta_p)$ is known as the evidence lower bound, which provides the lower bound of $\ln(p(\boldsymbol{y}|\theta_p))$. The idea of the EM approach is to maximize the lower bound $F(q,\theta_p)$ instead of maximizing the likelihood $p(\boldsymbol{y}|\theta_p)$.

The EM approach includes two steps. For instance, assuming the kth iteration:

- The expectation step: The parameter values from the last iteration are taken and assumed as constant in this step. And $p(y|\theta_{p,k-1})$ is also constant in this step. Therefore, $F(q,\theta_{p,k-1})$ could be maximized by minimizing $KL(q||p)$. The minimum value of $KL(q||p)$ is achieved if $KL(q||p)=0$, i.e.,

$$q(\boldsymbol{z})_k = p(\boldsymbol{z}|\boldsymbol{y};\theta_{p,k-1}). \tag{20}$$

  Moreover, the posterior distribution $p(\boldsymbol{z}|\boldsymbol{y};\theta_{p,k-1})$ can be analytically expressed with the Bayesian inference. Alternatively, the variational inference method with mean field approximation could also be implemented in the expectation step to approximate $q(\boldsymbol{z})_k$, if the derivation of the analytical expression of the posterior distribution is intractable.

- The maximization step: The obtained distribution of the latent variables $q(\boldsymbol{z})_k$ from the expectation step is taken and fixed in this step. And $F(q(\boldsymbol{z})_k,\theta_p)$ is maximized with manipulating the value of θ_p, which is

equivalent to the maximization problem:

$$\theta_{p,k} = \underset{\theta_p}{\operatorname{argmax}} \int q(\boldsymbol{z})_k \ln(p(\boldsymbol{y}, \boldsymbol{z}; \theta_p)) dz. \tag{21}$$

The integral term of the latent variables $\boldsymbol{z}$ in the objective function still exists. However, the natural logarithm function inside the integral function reduces its complexity and enables analytical solutions for the parameters θ (Bishop, 2006). Moreover, the additional distribution of $q(\boldsymbol{z})_k$ from the estimation step provides higher weight on the most likely latent variable sequence, which also alleviates the complexity of the optimization problem.

According to the EM approach with $KL(\cdot)$ as a non-negative measure, the relation between the likelihood functions from the $k-1$ and k iterations can be obtained. So a steady increment of the likelihood functions $p(\boldsymbol{y}|\theta_p)$ is observed and guaranteed in the EM approach.

$$\ln(p(\boldsymbol{y}|\theta_{p,k})) \geq F(q_k, \theta_{p,k}) \geq F(q_k, \theta_{k-1}) = \ln(p(\boldsymbol{y}|\theta_{p,k-1})) \tag{22}$$

The expectation and maximization steps are iterated until the terminal condition is fulfilled. For more details regarding the derivation, the proof, and the mathematical equations of the EM approach, the interested reader is referred to Bishop (2006) and Moon (1996). Moreover, to identify a meaningful number of Gaussian distributions N_C to build the GMD, the Bayesian information criterion (BIC), given in Eq. (23), is applied to determine the optimum number of component distributions (Bernardo et al., 2003; Gupta et al., 2010).

$$BIC = \ln(n_y) n_{\theta_p} - 2 \ln L(\boldsymbol{y}; \theta_p), \tag{23}$$

where n_y means the number of realizations, n_{θ_p} indicates the number of parameters, which is proportional to the number of component distributions. The configuration with the lowest BIC value is selected as the GMD of the original distribution. Here, the number of parameters are included in the objective function to avoid overfitting

Unlike the single Gaussian distribution and the nonlinear-transfer approach, the mean and the variance of $\boldsymbol{f}(\boldsymbol{\theta})$ are derived with the weighted sum of the mean and the variance of $\boldsymbol{f}(\boldsymbol{\theta})$ on the component distributions and are formulated as (Fisher, 1959; Bulmer, 1979):

$$\boldsymbol{\mu}_1(\boldsymbol{f}(\boldsymbol{\theta})) \approx \sum_{j=1}^{N_C} \omega_j \boldsymbol{\mu}_{1j}(\boldsymbol{f}(\boldsymbol{\theta})), \tag{24}$$

$$\boldsymbol{\mu}_2(\boldsymbol{f}(\boldsymbol{\theta})) \approx \left(\sum_{j=1}^{N_C} \omega_j (\boldsymbol{\mu}_{2j}(\boldsymbol{f}(\boldsymbol{\theta})) + \boldsymbol{\mu}_{1j}(\boldsymbol{f}(\boldsymbol{\theta}))^2) \right) - \boldsymbol{\mu}_1(\boldsymbol{f}(\boldsymbol{\theta}))^2, \tag{25}$$

where the mean and the variance of $\boldsymbol{f}(\boldsymbol{\theta})$ of the component distributions are calculated as:

$$\boldsymbol{\mu}_{1j}(\boldsymbol{f}(\boldsymbol{\theta})) \approx \sum_{i=1}^{2n_\theta^2+1} w_i \boldsymbol{f}(\theta_{ji}^{PEM}), \tag{26}$$

$$\boldsymbol{\mu}_{2j}(\boldsymbol{f}(\boldsymbol{\theta})) \approx \sum_{i=1}^{2n_\theta^2+1} w_i (\boldsymbol{f}(\theta_{ji}^{PEM}) - \boldsymbol{\mu}_{1j}(\boldsymbol{f}(\boldsymbol{\theta})))^2. \tag{27}$$

In summary, the proposed GMD-PEM approach retains most of the information of a non-Gaussian parameter distribution and does not introduce additional nonlinearities to the transformation step in Eq. (16) or in Eqs. (24) to (27). Note that the computational cost increases proportionally to the number of component distributions, and thus, we apply the BIC to ensure that only a low number of component distributions are used.

In the next section, we present the structure of a robust process design, especially the formulation of the objective function and constraints.

2.3. Robust process design with non-Gaussian uncertainties

The general structure of the probability-based robust process design reads as:

$$\min_{\mathbf{x}(\cdot),\mathbf{u}(\cdot)} \Phi(M(\mathbf{x}_{t_f})) \tag{28a}$$

subject to:

$$\dot{\mathbf{x}}_{\mathbf{d}}(t) = \mathbf{g}_{\mathbf{d}}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}), \tag{28b}$$

$$\mathbf{0} = \mathbf{g}_{\mathbf{a}}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}), \tag{28c}$$

$$\mathbf{x}_{\mathbf{d}}(0) = \mathbf{x}_0, \tag{28d}$$

$$\mathbf{Pr}[\mathbf{h}_{\mathbf{nq}}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}) \geq 0] \leq \varepsilon_{nq}, \tag{28e}$$

$$\mathbf{u}_{min} \leq \mathbf{u} \leq \mathbf{u}_{max}, \tag{28f}$$

where $t \in [0, t_f]$ is the time, $\mathbf{u} \in \mathbb{R}^{n_u}$ is the vector of the control variables, and $\mathbf{p} \in \mathbb{R}^{n_p}$ is the vector of the time-invariant parameters. $\mathbf{x}_d \in \mathbb{R}^{n_{x_d}}$ and $\mathbf{x}_a \in \mathbb{R}^{n_{x_a}}$ are the differential and algebra states; i.e., $\mathbf{x} = [\mathbf{x}_{\mathbf{d}}, \mathbf{x}_{\mathbf{a}}] \in \mathbb{R}^{n_x}$. The initial conditions for the differential states are given by $\mathbf{x}_0$, as uncertainties can exist in parameters and the initial conditions; i.e., $\boldsymbol{\theta} = [\mathbf{p}; \mathbf{x}_0]$. $\Phi(M(\mathbf{x}_{t_f}))$ denotes the robust formulation of the Mayer objective term $M(\mathbf{x}_{t_f})$ that is used for nominal optimal control problems. Eqs. (28b) and (28c) are the model equations with $\mathbf{g}_{\mathbf{d}} : \mathbb{R}^{(n_{x_d}+n_{x_a})\times n_u \times n_p} \rightarrow \mathbb{R}^{n_{x_d}}$ and $\mathbf{g}_{\mathbf{a}} : \mathbb{R}^{(n_{x_d}+n_{x_a})\times n_u \times n_p} \rightarrow \mathbb{R}^{n_{x_a}}$. The left side of Eq. (28e) is the probability of violating the inequality constraints $\mathbf{h}_{\mathbf{nq}} : \mathbb{R}^{(n_{x_d}+n_{x_a})\times n_u \times n_p} \rightarrow \mathbb{R}^{n_{nq}}$. ε_{nq} is the tolerance factor that gives the maximum acceptable probability for constraint violations. [$\mathbf{u}_{min}$,$\mathbf{u}_{max}$] are the upper and lower boundaries for the control variables.

$$\min_{\mathbf{x}(\cdot),\mathbf{u}(\cdot)} \quad \boldsymbol{\mu}_1(M(\mathbf{x}_{t_f})) + \alpha\boldsymbol{\mu}_2(M(\mathbf{x}_{t_f}))^{0.5}, \tag{29a}$$

subject to:

$$i = 1, \ldots, 2n_\theta^2 + 1 \quad j = 1, \ldots, N_C, \tag{29b}$$

$$\boldsymbol{\theta}_{ji} = [\mathbf{p}_{ji}, \mathbf{x}_{0,ji}]^T, \mathbf{x}_{ji} = [\mathbf{x}_{\mathbf{d},ji}, \mathbf{x}_{\mathbf{a},ji}]^T, \mathbf{x}_{\mathbf{d},ji}(0) = \mathbf{x}_{0,ji}, \mathbf{x}_{t_f,ji} = \mathbf{x}_{ji}(t_f), \tag{29c}$$

$$\dot{\mathbf{x}}_{\mathbf{d},ji}(t) = \mathbf{g}_{\mathbf{d}}(\mathbf{x}_{ji}(t), \mathbf{u}(t), \mathbf{p}_j i), \quad \mathbf{0} = \mathbf{g}_{\mathbf{a}}(\mathbf{x}_{ji}(t), \mathbf{u}(t), \mathbf{p}_{ji}), \tag{29d}$$

$$\overline{\boldsymbol{h}}_{ji} = -\mathbf{h}_{\mathbf{nq}}(\mathbf{x}_{ji}(t), \mathbf{u}(t), \mathbf{p}_{ji}) \tag{29e}$$

$$\sum_{j=1}^{N_C} \omega_j F\left(-\frac{\boldsymbol{\mu}_{1j}(\overline{\boldsymbol{h}})}{\boldsymbol{\mu}_{2j}(\overline{\boldsymbol{h}})^{0.5}}\right) \leq \varepsilon_{nq}, \tag{29f}$$

$$\boldsymbol{\mu}_1(M(\mathbf{x}_{t_f})) = \sum_{j=1}^{N_C} \omega_j \boldsymbol{\mu}_{1j}(M(\mathbf{x}_{t_f})), \tag{29g}$$

$$\boldsymbol{\mu}_2(M(\mathbf{x}_{t_f})) = \left(\sum_{j=1}^{N_C} \omega_j(\boldsymbol{\mu}_{2j}(M(\mathbf{x}_{t_f})) + \boldsymbol{\mu}_{1j}(M(\mathbf{x}_{t_f}))^2)\right) - \boldsymbol{\mu}_1(M(\mathbf{x}_{t_f}))^2, \tag{29h}$$

$$\boldsymbol{\mu}_{1j}(M(\mathbf{x}_{t_f})), \boldsymbol{\mu}_{2j}(M(\mathbf{x}_{t_f})), \boldsymbol{\mu}_{1j}(\overline{\boldsymbol{h}}), \boldsymbol{\mu}_{2j}(\overline{\boldsymbol{h}}) \text{are calculated with Eqs. (26) and (27)} \tag{29i}$$

$$\mathbf{u}_{min} \leq \mathbf{u} \leq \mathbf{u}_{max}, \tag{29j}$$

The proposed GMD-PEM algorithm provides meaningful results not only for non-Gaussian parameter uncertainties but also for non-Gaussian simulation results. The structure of the GMD-PEM algorithm for robust process design is shown in Eq. (29). Eq. (29a) is the robust objective function, which consists of the first and second statistical moment according to Eqs.(29g) and (29h). The weight factor α controls the trade-off between performance and robustness. Eqs. (29c) and (29d) are the model parameters, initial conditions and the equations of mathematical model. Here, i and j are the index of PEM sample points and component distribution, respectively. The failure probability of the inequality constraints is approximated with the weighted sum of the failure probability results of the individual component distributions as shown in Eq. (29f). For the sake of completeness, the robust process optimization problem making use of

the SGA and NTA concepts is summarized in Appendix A. Note that when the SGA or NTA is implemented, the approximation is accurate only if the probability distribution of the constraints is Gaussian. Thus, the SGA- and NTA-based approaches might lose essential information about the model output uncertainties in the case of non-Gaussian distributions, which might be the result of non-Gaussian model parameter uncertainties or due to model nonlinearities.

The performance of the GMD-PEM algorithm and its accuracy are compared with reference simulations in the next section.

3. Case study: the freeze-drying process

To demonstrate the performance of the GMD-PEM algorithm, we introduce the freeze-drying model first. Here, the model parameters are assumed to be uncertain and are assigned non-Gaussian distributions.

3.1. Primary drying of the freeze-drying process

The freeze-drying process, also known as lyophilization, has been extensively used in pharmaceutical manufacturing to stabilize APIs which are unstable and have limited storage time in aqueous solutions, e.g., therapeutic protein formulations and vaccines (Amorij et al., 2007; Kasper et al., 2013). However, the process suffers from high energy consumption, a long processing period, and critical operating limits (Fissore et al., 2011). The traditional freeze-drying process in a batch chamber consists of three successive steps which are freezing, primary drying, and secondary drying as illustrated in Fig. 3. The primary drying step is recognized as the most time-consuming and error-prone step (Fissore et al., 2010). Thus, in this study, we consider the robust process design of the primary drying step; i.e., we take parameter uncertainties into account while at the same time we aim to minimize the risk level of defective dried APIs.

Typically, during the primary drying step, the shelf temperature is increased, and the chamber pressure is decreased to a certain level so that sublimation of the water in the vials is initiated and continued at the sublimation surface,

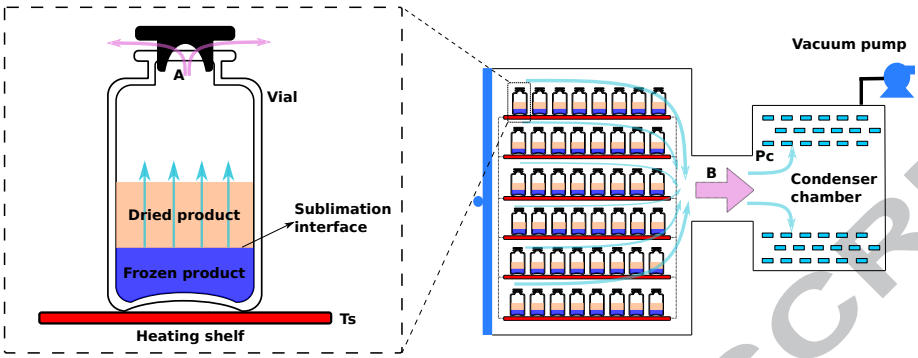


Figure 3: Schematic diagram of a lyophilizer where the temperature of the heating shelf (T_s) and the pressure of the chamber (P_c) are optimized to achieve maximum performance. On the left side is the enlarged view of the vial with the frozen and dried API. The light blue and red arrows indicate the flow of the vapor. A and B are the locations where a choked flow might exist.

which moves downward during the drying step. The water vapor is transferred from each vial to the condenser chamber and is discharged via the vacuum pump to keep the pressure in the chamber at a specific level. As the mass and energy transfer the occur in the vials is identical, the model of the primary drying step is based on a single vial as shown in Fig. 3. In this work, the mathematical model adapted from Mortier et al. (2016) is implemented for robust process design.

The mass transfer equation (Fissore et al., 2011), which describes the dynamics of the sublimation process at the sublimation surface, is given as:

$$\frac{\mathrm{d}m_{sub}}{\mathrm{d}t} = A_p \frac{P_i - P_c}{R_p}, \tag{30}$$

where A_p is the cross-sectional area of the product, P_c is the chamber pressure, and R_p is the dried product resistance to the vapor flux. P_i is the vapor pressure at the sublimation interface which can be calculated according to the equation in Murphy & Koop (2005):

$$P_i = \exp(9.55 - \frac{5720}{T_i} + 3.53\ln(T_i) - 0.00728T_i). \tag{31}$$

T_i is the temperature at the sublimation interface and is calculated with the

energy balance equation given in Fissore et al. (2011). Here, the heat used for sublimation is assumed to be equal to the heat transferred from the heating shelf:

$$K_v(T_s - T_B)A_v = \Delta H_s \frac{\mathrm{d}m_{sub}}{\mathrm{d}t}, \tag{32}$$

where K_v, A_v, and T_s denote the heat transfer coefficient, the outer cross-sectional area of the vial, and the shelf temperature, respectively. ΔH_s is the heat of sublimation as given in Murphy & Koop (2005) and reads as:

$$\Delta H_s = 4.68 \times 10^4 + 35.9T_i - 0.0741T_i^2 + 542\exp(-(\frac{T_i}{124})^2). \tag{33}$$

T_B is the temperature at the bottom of vial and is equal to

$$T_B = T_i + \Delta T, \tag{34}$$

where ΔT is the temperature difference across the frozen layer and is calculated with the following equation given by Mortier et al. (2016):

$$\Delta T = \frac{889200\frac{(L_f(P_i - P_c))}{R_p} - 0.0102L_f(T_s - T_i)}{1 - 0.0102L_f}. \tag{35}$$

Here, L_f is the height of the frozen layer and has the following relation with m_{sub}:

$$m_{sub} = (L_{total} - L_f)\rho_I \epsilon A_p, \tag{36}$$

where L_{total}, ρ_I, and ϵ are the total height of the product layer, the density of the ice, and the volume of the ice fraction, respectively. The model parameters and the size of the vials, are taken from Mortier et al. (2016). Note that the value of R_p and K_v in Table 1 are the averages of their values at different conditions. More details regarding the determination, the structure, and coefficients R_p and K_v can be found in Pisano et al. (2013) and Mortier et al. (2016).

The mathematical model is used in this case study to maximize the efficiency of the primary drying step under parameter uncertainties in R_p and K_v while ensuring the product quality at the same time. Thus, the objective function is to maximize the total mass of the ice removed by sublimation and to minimize the operating time. To avoid irreversible product damage and have an

Table 1: Nominal values of the model parameters and the initial conditions for the primary drying model.

Parameters	Unit	Nominal Value
A_p	m^2	3.80×10^{-4}
A_v	m^2	4.15×10^{-4}
$A_{v,n}$	m^2	1.25×10^{-4}
R_p	m/s	5.57×10^4
K_v	$J/(m^2sK)$	11.47
L_{total}	m	0.00658
ρ_I	kg/m^3	919
ϵ	–	0.97
M	kg/mol	0.018
k	–	1.33
R	J/(Kmol)	8.314
T_g	°C	-34

acceptable API cake appearance, the production temperature at the sublimation interface should be carefully maintained below the critical collapse temperature T_c, which in this case is assumed to be equal to the glass transition temperature T_g (Mortier et al., 2016). Additionally, an upper boundary is given for the sublimation rate $\frac{\mathrm{d}m_{sub}}{\mathrm{d}t}$, which is due to the choke flow phenomenon at the vial neck as explained in Searles (2004). The sublimation rate is calculated using the following equation as in Mortier et al. (2016):

$$\dot{m}_{sub,choke,vial} = \frac{0.3A_{v,n}\sqrt{\frac{kT_rR}{M}}M}{RT_r}P_c, \tag{37}$$

where $A_{v,n}$ is the cross area at the vial neck. Chamber pressure P_c and shelf temperature T_s are manipulated within the range from [5 Pa 30 Pa] and [-40 °C 30 °C], respectively. Note that the lower boundary of P_c is normally not set below 5 bar, as a very low chamber pressure may have problems with product contamination and heterogeneous heat transfer (Tang & Pikal, 2004). The un-

certainty of the two parameters is assumed to follow a non-Gaussian distribution and is characterized by the samples in the scatter plot in Fig. 4. Please note that the assumption of the parameter uncertainties is based on values provided in Mortier et al. (2016) and the reference therein. To demonstrate the performance of SGA, NTA, and GMD-PEM, artificial samples are created based on the assumed non-Gaussian distribution. The result structure of the nominal optimization of the primary drying process is given in Appendix B. Moreover, the structure of the robust optimization of it can be straightforwardly derived with Eq. 29 and the equations in Appendix A.

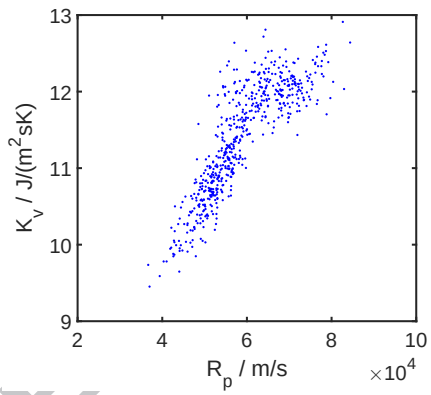


Figure 4: Scatter plot of the sample points for two uncertain parameters R_p and K_v.

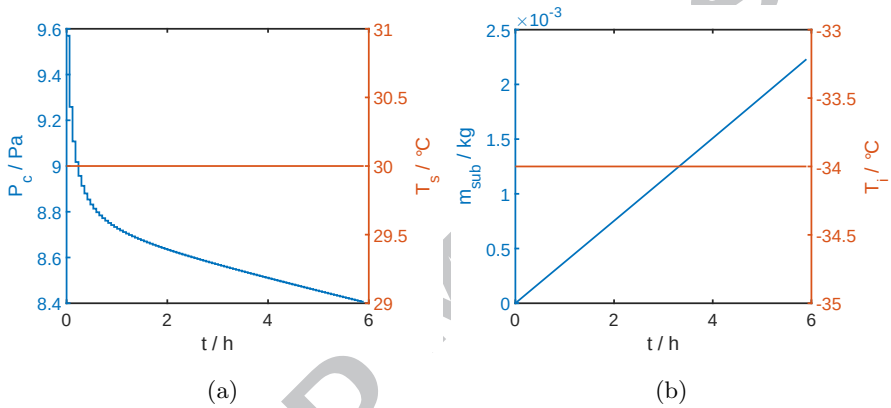
The case study is coded in MATLAB®. The robust process design was solved with the simultaneous approach (Biegler, 2007) which was implemented in the symbolic framework CasADi for numerical optimization (Andersson et al., 2012) using the NLP solver IPOPT (Wächter & Biegler, 2006) and the MA57 linear solver (Duff, 2004). The EM algorithm is initialized by k-means (Krishna & Murty, 1999).

4. Results and discussion

This section discusses the robust process design results of the primary freeze-drying step. First, the results for the nominal process design are given. Next,

the adverse effect of the parameter uncertainties on the nominal process design is presented. Alternatively, the usefulness of the robust process design using the GMD-PEM algorithm is introduced. To this end, the accuracy of the approximated statistical moments and distributions are analyzed, and the results of the robust process design for the freeze-drying step are compared and discussed with state-of-the-art methods in the field of robust process design.

4.1. Results for the nominal process design



(a) (b)

Figure 5: Control profile of chamber pressure P_c and shelf temperature T_s (a), and evolution profile of the mass of the ice removed by sublimation and the temperature at the sublimation interface (b) from the nominal design.

First, we want to optimize the efficiency of the primary drying process in the absence of parameter uncertainties. As we described in the last section, shelf temperature T_s and chamber pressure P_c are designed to achieve the maximum amount of ice removed by sublimation within the shortest drying time, while the appearance of the dried product is guaranteed by limiting temperature T_i at the sublimation interface below critical temperature T_c. The designed temperature and pressure profiles are shown in Fig. 5. As in Fig. 5a, T_s is kept at its upper boundary, i.e., 30 °C, to provide more energy for the sublimation. In Fig. 5b, T_i is also kept at its upper boundary to ensure higher vapor pressure P_i at the sublimation interface to accelerate the sublimation process according to Eqs. 30

and 31. In the beginning, P_c is set to 9.6 bar to achieve a higher sublimation speed and is decreased gradually to compensate the influence of the decreasing height of the frozen layer in accordance with Eq. 35. Note that the complete expressions of R_p and K_v from Mortier et al. (2016) are used for the nominal process design. Note that Mortier et al. (2016) also attempted to optimize the primary drying in the freeze-drying process with a grid-based approach. In contrast to our results, the grid-based design compensates for the influence of the decreasing height of the frozen layer by decreasing the shelf temperature, which leads to a certain loss in the sublimation speed. Consequently, we need less than 6 h rather than the almost 7 h in Mortier et al. (2016) to complete the primary drying process as indicated by the curve of m_{sub} in Fig. 5b. Note that the transition phase at the beginning, i.e., heating the shelf and vacuuming the chamber to the design value, is neglected in this work because the period is quite short compared to the entire primary drying step. In addition, the choked flow limit at the vial neck calculated with Eq. 37 is not activated because the sublimation speed is far below the limit, which is also shown in the following subsection.

4.2. Effect of parameter uncertainties on the nominal process design

The optimal process design for the nominal case above is based on the assumption that the model parameters are accurate. However, due to measurement imperfections and model simplifications, model parameters derived from noisy measurement data are imprecise and might be described best via arbitrarily distributed random variables (Rossner, 2014). In this particular case, we assume the two parameters R_p and K_v are uncertain as indicated in the scatter plot in Fig. 4. Next, we analyze the adverse effect of these parameter uncertainties on the performance of the primary drying process obtained by ignoring these parameter uncertainties. The results are illustrated in Fig. 6. As can be observed, the curves for the evolution of m_{sub} and T_i in Figs. 6a and 6c deviate from the nominal values and vary in certain ranges expressed with the confidence intervals (CIs) (Fisher, 1959). The primary drying process

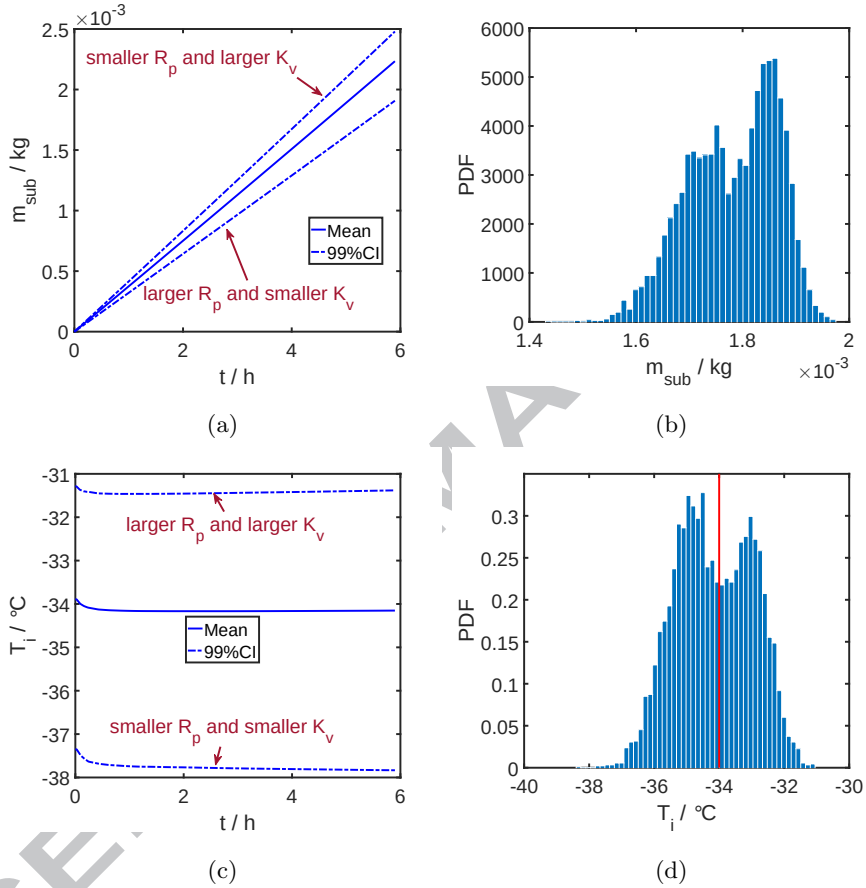
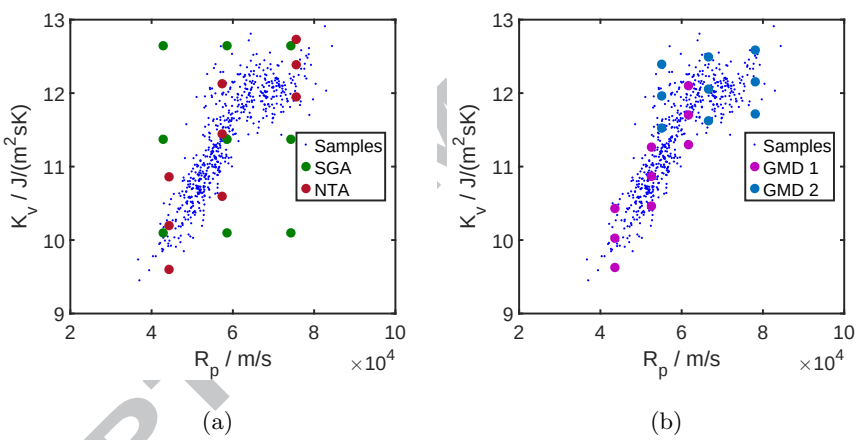


Figure 6: Evolution of the mean and 99% confidence interval (CI) of the mass of the ice removed by sublimation m_{sub} (a) and the temperature at sublimation interface T_i (c). (b) and (d) are the plots of the probability density function (PDF) of m_{sub} and T_i at t = 5 h, respectively.

with smaller (higher) values for R_p and K_v has a lower (higher) T_i. In contrast, with smaller R_p and larger K_v values, the efficiency of the process is higher than what we expected regarding the nominal process design. As T_i exceeds the critical temperature $T_c = -34°\mathrm{C}$, this leads to an undesired collapse of the API cake. In Fig. 6d, there is a high risk that the API product will be wasted as almost half of the probability distribution is on the right side of the red line, i.e., above the critical temperature. Therefore, it is necessary to consider parameter uncertainties in the design of the primary drying process as part of the robust process design.

4.3. Deterministic samples



(a) (b)

Figure 7: Illustration of the original samples and the deterministic samples derived from the three approaches. The green and red points in (a) are the deterministic samples from the single Gaussian approach (SGA) and the nonlinear-transfer approach (NTA), respectively. The purple and blue points in (b) are the deterministic samples for two component distributions of the Gaussian mixture distribution (GMD).

As discussed in Section 2, the GMD-PEM algorithm can be used to propagate and to quantify uncertainties in the field of robust process design efficiently. The resulting samples are plotted in Fig. 7 in comparison with sample points derived with the SGA and the NTA. The green sample points in Fig. 7a are derived with the SGA. They are symmetrically distributed in the parameter

space as only the mean and the variance of the individual parameters are used; see Section 2.1. The red sample points in Fig. 7a are derived with the NTA. The NTA sample points are asymmetrically located in the parameter space due to the nonlinear transformation step addressing the correlation and the non-Gaussian shape of the parameter distributions. The EM algorithm and the BIC as given in Section 2.1 are used to determine the number and the coefficients of the Gaussian component distributions. In Fig. 7b, two component distributions are illustrated which have the best match with the parameter uncertainties. The corresponding deterministic samples are highlighted individually in Fig. 7b with purple and blue sample points, respectively. Although the number of GMD-PEM samples is twice as many as those for SGA or NTA, they can represent the shape of the original parameter uncertainties more accurately with lower computational cost compared to Monte Carlo simulations.

4.4. Approximation accuracy

In this section, we analyze the accuracy of the novel GMD-PEM algorithm regarding the approximated mean and the variance of the freeze-drying process relevant state variables, i.e., m_{sub} and T_i. The results are summarized in Fig. 8. Here, the approximated mean and variance values are compared with the references from the Monte Carlo simulations. As shown in Figs. 8a and 8b, the mean of m_{sub} can be estimated accurately with the SGA, NTA, and GMD-PEM algorithm, while the variances are estimated more accurately with the NTA and the GMD-PEM algorithm. Note that the approximated variances derived with the NTA deviate slightly from the reference, but the deviation is considerably smaller compared to the estimation based on the SGA; see Fig. 8b. The same analysis for T_i is illustrated in Figs. 8c and 8d. The estimated means and variances from the GMD-PEM algorithm are in good accordance with the references. The estimated mean based on the SGA or NTA is slightly lower than the references. However, the estimated variance deviates considerably from the reference, especially the variance approximated by the SGA. The GMD-PEM algorithm, in turn, not only provides a more accurate estimate of the

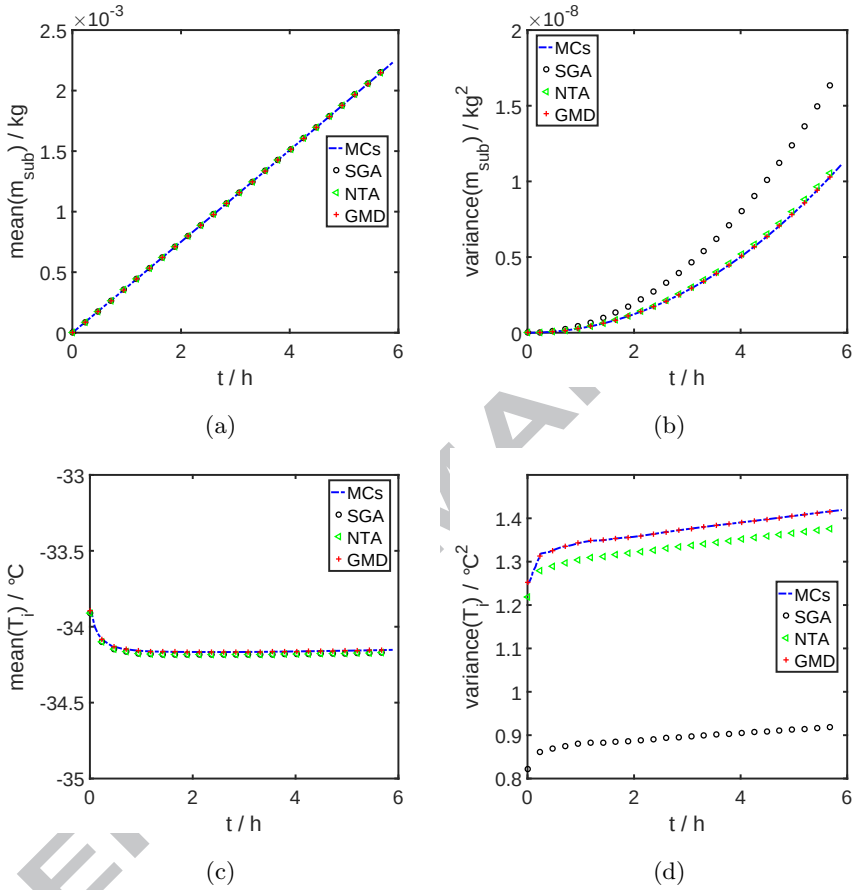
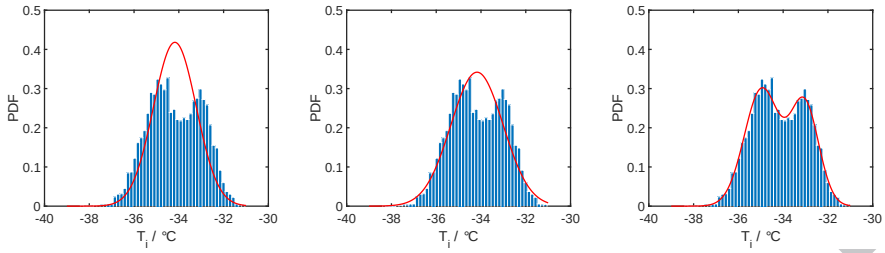


Figure 8: Comparison of the mean and the variance of the mass of the ice removed by sublimation m_{sub} and the temperature at sublimation interface T_i, which are estimated with the single Gaussian approach (SGA), the nonlinear-transfer approach (NTA), and the Gaussian mixture distribution (GMD). Values from Monte Carlo simulations (MCs) are used as references.



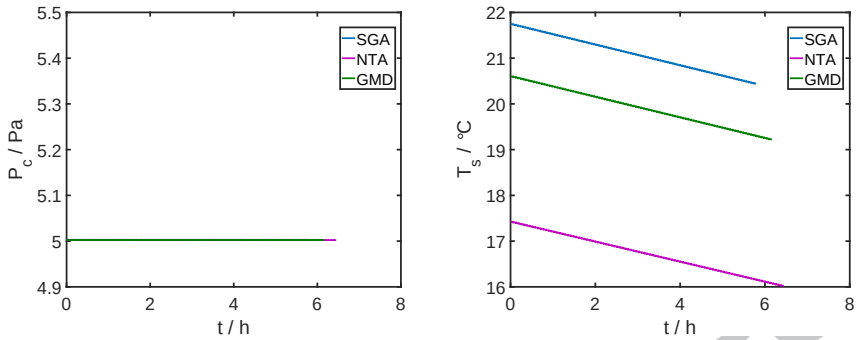
(a) SGA (b) NTA (c) GMD

Figure 9: Comparison of the original probability density function of the temperature at sublimation interface T_i at a single time point with the ones approximated by the single Gaussian approach (SGA), the nonlinear-transfer approach (NTA), and the Gaussian mixture distribution (GMD).

mean and the variance but also represents the non-Gaussian shape of the model output distributions which is essential for approximating the robust inequality constraints given in Eq. (29f). For instance, the non-Gaussian distribution can be observed from the temperature at the sublimation interface T_i given in Fig. 9.

4.5. Robust process design results

The robust process design with the structure given in Eqs. A.1 and 29 is implemented with the different samples in Fig. 7 to design the primary drying process. The tolerance factor ε_{nq} is set to 1%, so that the risk of failure in the API cake appearance is lower than 1%. The designed profiles for the control variables P_c and T_s are compared in Fig. 10. As we can see, the optimal shelf temperature is lower than the upper boundary to force T_i to be lower than the critical collapse temperature with 99% probability. In parallel, the chamber pressure decreases to its lower boundary 5 bar to accelerate the sublimation. While the chamber pressure is fixed, T_s is gradually decreased to compensate for the influence of decreasing height of the frozen layer. As discussed in Section 4.4, the approximation accuracy of the SGA, NTA, and GMD-PEM is different, and thus, the optimized T_s is different as well. To compare the performance of



(a) (b)

Figure 10: Results from a robust process design with the deterministic samples from the single Gaussian approach (SGA), the nonlinear-transfer approach (NTA), and the Gaussian mixture distribution (GMD). (a) and (b) are the profiles for chamber pressure P_c and shelf temperature T_s, respectively.

the results obtained from the SGA, NTA, and GMD-PEM algorithm, we simulate the primary drying process with 6000 samples that are derived from the original parameter distributions; see Fig. 4. The mean value of m_{sub}, i.e., $\overline{m}_{sub}$, at the final time point and the probability of constraint violations (ϵ_v) are calculated based on the simulation results which are related to the 6000 parameter samples. The results are listed in Table 2. The nominal design has the best efficiency: a low freeze-drying time and high m_{sub}. However, almost 50% of the simulations violate the temperature constraint; i.e., T_i exceeds the critical temperature. The SGA result, in turn, has a better performance compared to the results from the NTA and the GMD-PEM algorithm, but the constraint violation is $\frac{94}{6000} = 1.6\%$ which is almost twice as high as the target value $\varepsilon_{nq} = 1\%$. The resulting NTA setting has much fewer constraint violations, but the efficiency is low with 6.45 h for the primary drying process. In other words, the design is too conservative. In contrast, the derived robust design from the GMD-PEM algorithm has the best trade-off between process efficiency and product quality, i.e., the API cake appearance and integrity. In Table 2, the GMD-PEM design achieves the same m_{sub} within 6 h and results in a failure

probability of $\frac{62}{6000} = 1\%$. Moreover, computational costs, i.e., the number of model evaluations for uncertainty quantification for each optimization iteration, are also given in Table 2. As can be seen, GMD has the highest computational costa as a single non-Gaussian distribution is approximated with several Gaussian distributions, and for each of these Gaussian distributions, PEM samples are generated. Therefore, the proposed GMD-PEM concept for the robust process design has better performance at the cost of higher computational expense. Please note that the computational cost of GMD-PEM increases proportionally with the increasing number of component distributions. This is also the reason why BIC is used to select the optimum number of component distributions. The

Table 2: Results from the nominal design and the robust design with the deterministic samples from the single Gaussian approach (SGA), the nonlinear-transfer approach (NTA), and the Gaussian mixture distribution (GMD). 6000 samples generated from the original samples are used to validate the probability of constraint violations (ϵ_v) for the different methods. Computational cost indicates the number of model evaluations that are needed to quantify the uncertainties for each optimization iteration.

	drying time [h]	$\overline{m}_{sub} \times 10^3$ [kg]	ϵ_v	comput. cost [-]
Nominal	5.90	2.2	50%	1
SGA	5.80	2.0	1.6%	9
NTA	6.45	2.0	0.15%	9
GMD	5.98	2.0	1%	18

choked flow limit is investigated for the GMD-PEM design, and the sublimation speed is considerably slower than the choked flow limit, as shown in Fig. 11. Thus, the flow rate of the vapor at the vial neck is not fast enough to reach the speed of sound and to trigger the choked flow phenomenon (Searles, 2004).

5. Conclusions

In this paper, a novel GMD-PEM algorithm for a robust process design with the parameter uncertainties of non-Gaussian distributions was investigated and

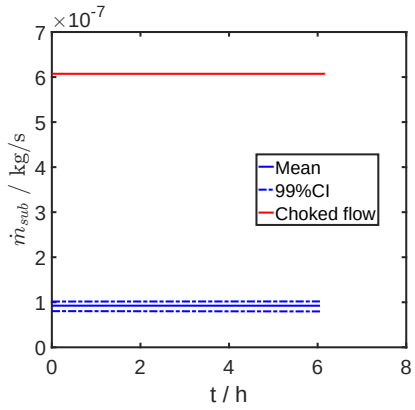


Figure 11: Comparison of the sublimation rates during the primary drying process and the choked flow limitation.

successfully demonstrated for the primary drying process. The GMD-PEM algorithm was benchmarked with two approaches commonly used in literature and introduced in our previous work. The proposed robust process design is based on Gaussian mixture distributions and the expectation-maximization algorithm. Deterministic parameter samples were generated and used to approximate the mean and the variance of the objective function. In parallel, also under parameter uncertainties, the temperature at the sublimation interface had to be kept reliably below -34 °C to ensure high-quality products. First, however, the primary drying step was optimized while the parameter uncertainties were ignored. A design with high efficiency was obtained compared to the literature but might result in low-quality products in practice due to the neglected parameter uncertainties. Second, the GMD-PEM algorithm was implemented and based on the simulation results was able to calculate the most reliably robust process design for the freeze-drying process. As demonstrated, the proposed algorithm can approximate the non-Gaussian model parameters and the output distributions adequately, which is essential in fulfilling inequality constraints under non-Gaussian parameter uncertainties. Monte Carlo simulations were used to evaluate the final design. In comparison to standard PEM concepts, the GMD-PEM algorithm provides the best trade-off between process efficiency and

the acceptable limit of constraint violations of the product-relevant sublimation interface temperature.

Nomenclature

Methodology

$\boldsymbol{\mu}_{1j}$	mean of the j-th component distribution
$\boldsymbol{\Sigma}$	covariance matrix of the random parameters $\boldsymbol{\xi}$
$\boldsymbol{\Sigma}_i$	covariance matrix of the j-th component distribution
$\boldsymbol{\theta}$	random parameters with non-Gaussian distribution
$\boldsymbol{\theta}_i^{PEM}$	i-th sample point for random variable $\boldsymbol{\theta}$ from PEM and SGA or NTA
$\boldsymbol{\theta}_{ji}^{PEM}$	i-th sample point from PEM of the random variables $\boldsymbol{\theta}$ from the j-th component distribution
$\boldsymbol{\xi}$	random parameters with Gaussian distribution
$\boldsymbol{\xi}_i^{PEM}$	i-th sample point for random variable $\boldsymbol{\xi}$ from PEM and SGA or NTA
$\boldsymbol{f}(\cdot)$	nonlinear functions
$\boldsymbol{p}(\boldsymbol{\theta})$	probability density function of the random variables $\boldsymbol{\theta}$
$\boldsymbol{p}(\boldsymbol{\xi})$	joint density function of random parameters $\boldsymbol{\xi}$
$\boldsymbol{p}_j(\boldsymbol{\theta})$	probability density function of j-th component distribution
$\boldsymbol{x}$	model state variables
$\boldsymbol{y}$	realizations from the non-Gaussian distributions
$\mathbb{E}(\boldsymbol{\theta})$	mean value of the random parameters $\boldsymbol{\theta}$
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	the Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$\mu_{\boldsymbol{\xi}}$	mean values of the random parameters $\boldsymbol{\xi}$

$\mu_{\boldsymbol{n}}$	n-th statistical moment of $\boldsymbol{x}$
ω_j	ω_j is the non-negative weight for j-th component distribution
$\Sigma(\boldsymbol{\theta})$	covariance matrix of the random parameters $\boldsymbol{\theta}$
Σ_ρ	correlation matrix
θ_p	unknown parameters of the component distributions
BIC	Baysian information criterion
F	CDF of univariate Gaussian distribution
$F(q, \theta_p)$	evidence lower bound
F_i	marginal CDF of parameter θ_i, $i = 1, \ldots, n_\theta$
F_N	joint CDF of multivariate Gaussian distribution
$F_{\boldsymbol{\theta}}$	joint CDF for the random parameter $\boldsymbol{\theta}$
F_ξ	joint CDF for the random parameter $\boldsymbol{\xi}$
$GF[\cdot]$	generator function
h_{nq}	inequality constraint
i	index for sample point from PEM, $i = 1, \ldots, 2n_\theta^2 + 1$
j	index for component distribution, $j = 1, \ldots, NC$
k	iteration number in the EM approach
$KL(\cdot)$	Kullback-Leibler divergence
$L(\cdot)$	likelihood function
$M(\cdot)$	objective function of the optimization problem
n_y	number of realizations $\boldsymbol{y}$
n_{θ_p}	number of unknown parameters θ_p

n_θ	dimension of the random variable $\boldsymbol{\theta}$	
n_ξ	dimension of the random parameters $\boldsymbol{\xi}$	
NC	number of the component distributions	
PEM	point estimate method	
$q(\boldsymbol{z})$	probability distribution of latent variables $\boldsymbol{z}$	
w_0, w_1, w_2	weights for generator functions in PEM	
z	latent variables	
CDF	cumulative density function	
EM	expectation and maximization approach	
GMD	Gaussiand mixture distribution	
NTA	nonlinear-transfer approach	
SGA	single Gaussian approach	
Case study		
ΔH_s	heat of sublimation,	J/kg
$\dot{m}_{sub,choke,vial}$	choke flow limited sublimation rate,	kg/s
ϵ	volume of the ice fraction,	$-$
ρ_I	density of the ice,	kg/m^3
A_p	cross-sectional area of the product,	m^2
A_v	outer cross-sectional area of the vial,	m^2
$A_{v,n}$	cross area at the vial neck,	m^2
K_v	heat transfer coefficient,	$J/(m^2sK)$
L_f	height of the frozen layer,	m

L_{total}	total height of the product layer,	m
P_c	chamber pressure,	Pa
P_i	vapor pressure at the sublimation interface,	Pa
R_p	dried product resistance to the vapor flux,	m/s
T_B	temperature at the bottom of vial,	K
T_i	temperature at the sublimation interface,	K
T_s	shelf temperature,	K
m_{sub}	mass of the sublimated ices, ,	kg

Appendix A. Robust process design with the SGA and the NTA

Here, we present the structure of the robust process design where the single Gaussian approach and the nonlinear-transfer approach are implemented. As mentioned, only the first two statistical moments can be estimated through the PEM with good accuracy. Thus, only the mean and the variance of the Mayer objective term and the inequality constraints are used to approximate their robust formulation in Eqs. (28a) and (28e). The resulting robust optimization structure reads as:

$$\min_{\mathbf{x}(\cdot),\mathbf{u}(\cdot)} \ \boldsymbol{\mu}_1(M(\mathbf{x}_{t_f})) + \alpha\boldsymbol{\mu}_2(M(\mathbf{x}_{t_f}))^{0.5}, \tag{A.1a}$$

subject to:

$$i = 1, \ldots, 2n_\theta^2 + 1 \tag{A.1b}$$

$$\boldsymbol{\theta}_i = [\mathbf{p}_i, \mathbf{x}_{0,i}]^T, \mathbf{x}_i = [\mathbf{x}_{\mathbf{d},i}, \mathbf{x}_{\mathbf{a},i}]^T, \mathbf{x}_{\mathbf{d},i}(0) = \mathbf{x}_{0,i}, \mathbf{x}_{t_f,i} = \mathbf{x}_i(t_f), \tag{A.1c}$$

$$\dot{\mathbf{x}}_{\mathbf{d},i}(t) = \mathbf{g}_{\mathbf{d}}(\mathbf{x}_i(t), \mathbf{u}(t), \mathbf{p}_i), \quad \mathbf{0} = \mathbf{g}_{\mathbf{a}}(\mathbf{x}_i(t), \mathbf{u}(t), \mathbf{p}_i), \tag{A.1d}$$

$$\overline{\boldsymbol{h}}_i = -\mathbf{h}_{\mathbf{nq}}(\mathbf{x}_i(t), \mathbf{u}(t), \mathbf{p}_i) \tag{A.1e}$$

$$F\left(-\frac{\boldsymbol{\mu}_1(\overline{\boldsymbol{h}})}{\boldsymbol{\mu}_2(\overline{\boldsymbol{h}})^{0.5}}\right) \leq \varepsilon_{nq}, \tag{A.1f}$$

$$\boldsymbol{\mu}_1(M(\mathbf{x}_{t_f})) = \sum_{i=1}^{2n_\theta^2+1} w_i M(\mathbf{x}_{t_{f,i}}), \tag{A.1g}$$

$$\boldsymbol{\mu}_2(M(\mathbf{x}_{t_f})) = \sum_{i=1}^{2n_\theta^2+1} w_i (M(\mathbf{x}_{t_{f,i}}) - \boldsymbol{\mu}_1(M(\mathbf{x}_{t_f})))^2, \tag{A.1h}$$

$$\boldsymbol{\mu}_1(\overline{\boldsymbol{h}}) = \sum_{i=1}^{2n_\theta^2+1} w_i \overline{\boldsymbol{h}}_i, \tag{A.1i}$$

$$\boldsymbol{\mu}_2(\overline{\boldsymbol{h}}) = \sum_{i=1}^{2n_\theta^2+1} w_i (\overline{\boldsymbol{h}}_i - \boldsymbol{\mu}_1(\overline{\boldsymbol{h}}))^2, \tag{A.1j}$$

$$\mathbf{u}_{min} \leq \mathbf{u} \leq \mathbf{u}_{max}, \tag{A.1k}$$

where $\boldsymbol{\theta}_i$ are the deterministic PEM samples. Weight factor α controls the process performance and the robustness. Eq. (A.1f) approximates the failure probability of the inequality constraints (Zhao & Ono, 2001).

Appendix B. Structure of nominal optimization of the primary drying process

The structure of the nominal optimization of the primary drying process is given below. Chamber pressure P_c and shelf temperature T_s are manipulated to minimize final drying time t_f (B.1a), with which the frozen product can be completely dried (B.1g), and satisfy the constraints for the CQA (B.1e) and the technical limitation (B.1f):

$$\min_{\mathbf{T_i}(\cdot), \mathbf{m_{sub}}(\cdot), \mathbf{T_s}(\cdot), \mathbf{P_c}(\cdot)} t_f, \tag{B.1a}$$

subject to:

Mathemical model:	Eqs.(30) to (36)	(B.1b)
Bounds:	$P_c^L \leq P_c(t) \leq P_c^U$	(B.1c)
	$T_s^L \leq T_s(t) \leq T_s^U$	(B.1d)
Inequality constraints:	$T_i(t) \leq T_c$	(B.1e)
	$\dot{m}_{sub}(t) \leq \dot{m}_{sub,choke,vial}$	(B.1f)
Equality constraints:	$m_{sub}(t_f) = \rho_I \epsilon A_p L_{total}$	(B.1g)

Acknowledgement

X. Xie acknowledges funding from the "Promotionsprogramm μ-Props" by MWK Niedersachsen and the support from the International Max Planck Research School for Advanced Methods in Process and Systems Engineering, MPI Magdeburg.

References

Amorij, J., Meulenaar, J., Hinrichs, W., Stegmann, T., Huckriede, A., Coenen, F., & Frijlink, H. (2007). Rational design of an influenza subunit vaccine powder with sugar glass technology: preventing conformational changes of haemagglutinin during freezing and freeze-drying. *Vaccine*, *25*, 6447–6457.

Andersson, J., Åkesson, J., & Diehl, M. (2012). Casadi: A symbolic package for automatic differentiation and optimal control. *Recent Advances in Algorithmic Differentiation*, *87*, 297–307.

Benyahia, B., Lakerveld, R., & Barton, P. I. (2012). A plant-wide dynamic model of a continuous pharmaceutical process. *Industrial & Engineering Chemistry Eesearch*, *51*, 15393–15412.

Bernardo, J., Bayarri, M., Berger, J., Dawid, A., Heckerman, D., Smith, A., West, M. et al. (2003). The variational bayesian em algorithm for incomplete data: with application to scoring graphical model structures. *Bayesian Statistics*, *7*, 453–464.

Beyer, H.-G., & Sendhoff, B. (2007). Robust optimization–a comprehensive survey. *Computer Methods in Applied Mechanics and Engineering*, *196*, 3190–3218.

Biegler, L. T. (2007). An overview of simultaneous strategies for dynamic optimization. *Chemical Engineering and Processing: Process Intensification*, *46*, 1043–1053.

Biegler, L. T. (2010). *Nonlinear programming: concepts, algorithms, and applications to chemical processes*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics.

Bishop, C. (2006). *Pattern recognition and machine learning*. Berlin, Heidelberg: Springer-Verlag.

Botev, Z. I., Grotowski, J. F., Kroese, D. P. et al. (2010). Kernel density estimation via diffusion. *The Annals of Statistics*, *38*, 2916–2957.

Bulmer, M. G. (1979). *Principles of statistics*. Oxford, UK: Courier Corporation.

Carullo, A., & Vallan, A. (2012). Measurement uncertainty issues in freeze-drying processes. *Measurement*, *45*, 1706–1712.

Duff, I. S. (2004). Ma57—a code for the solution of sparse symmetric definite and indefinite systems. *ACM Transactions on Mathematical Software (TOMS)*, *30*, 118–144.

Edgar, T. F., Himmelblau, D. M., Lasdon, L. S. et al. (2001). *Optimization of chemical processes*. New York, NY, USA: McGraw-Hill.

Emery, A., & Nenarokomov, A. V. (1998). Optimal experiment design. *Measurement Science and Technology*, *9*, 864.

Fisher, R. A. (1959). *Statistical methods and scientific inference*. (2nd ed.). Edinburgh: Oliver & Boyd.

Fissore, D., Pisano, R., & Barresi, A. A. (2010). On the methods based on the pressure rise test for monitoring a freeze-drying process. *Drying Technology*, *29*, 73–90.

Fissore, D., Pisano, R., & Barresi, A. A. (2011). Advanced approach to build the design space for the primary drying of a pharmaceutical freeze-drying process. *Journal of Pharmaceutical Sciences*, *100*, 4922–4933.

Grossmann, I., & Sargent, R. (1978). Optimum design of chemical plants with uncertain parameters. *AIChE Journal*, *24*, 1021–1028.

Gupta, U. D., Menon, V., & Babbar, U. (2010). Parameter selection for em clustering using information criterion and pddp. *International Journal of Engineering and Technology*, *2*, 340.

ICH (2009). Pharmaceutical development q8. *ICH Harmon Tripart Guidel*, *8*, 1–28.

Joshi, M., Seidel-Morgenstern, A., & Kremling, A. (2006). Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. *Metabolic Engineering*, *8*, 447–455.

Kaiser, N. M., Flassig, R. J., & Sundmacher, K. (2016). Probabilistic reactor design in the framework of elementary process functions. *Computers & Chemical Engineering*, *94*, 45–59.

Kalyanaraman, J., Fan, Y., Labreche, Y., Lively, R. P., Kawajiri, Y., & Realff, M. J. (2015). Bayesian estimation of parametric uncertainties, quantification and reduction using optimal design of experiments for co 2 adsorption on amine sorbents. *Computers & Chemical Engineering*, *81*, 376–388.

Kasper, J. C., Winter, G., & Friess, W. (2013). Recent advances and further challenges in lyophilization. *European Journal of Pharmaceutics and Biopharmaceutics*, *85*, 162–169.

Kleinbaum, D. G., & Klein, M. (2010). *Maximum likelihood techniques: An overview*. Berlin, Heidelberg: Springer-Verlag.

Krishna, K., & Murty, M. N. (1999). Genetic k-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *29*, 433–439.

Lebrun, R., & Dutfoy, A. (2009). A generalization of the nataf transformation to distributions with elliptical copula. *Probabilistic Engineering Mechanics*, *24*, 172–178.

Lerner, U. N. (2002). *Hybrid Bayesian networks for reasoning about complex systems*. Ph.D. thesis ,Stanford University, Stanford.

Liu, Y., & Xu, X. (2017). Industry 4.0 and cloud manufacturing: A comparative analysis. *Journal of Manufacturing Science and Engineering*, *139*, 034701.

Luyben, W. L. (1989). *Process modeling, simulation and control for chemical engineers*. New York, NY, USA: McGraw-Hill.

Maußner, J., & Freund, H. (2018). Optimization under uncertainty in chemical engineering: Comparative evaluation of unscented transformation methods and cubature rules. *Chemical Engineering Science*, *183*, 329–345.

Mesbah, A., Streif, S., Findeisen, R., & Braatz, R. D. (2014). Stochastic nonlinear model predictive control with probabilistic constraints. *2014 American Control Conference (ACC)*, *1*, 2413–2419.

Moon, T. K. (1996). The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, *13*, 47–60.

Mooney, C. Z. (1997). *Monte Carlo Simulation*. Thousand Oaks, CA: SAGE Publications.

Mortier, S. T. F., Van Bockstal, P.-J., Corver, J., Nopens, I., Gernaey, K. V., & De Beer, T. (2016). Uncertainty analysis as essential step in the establishment of the dynamic design space of primary drying during freeze-drying. *European Journal of Pharmaceutics and Biopharmaceutics*, *103*, 71–83.

Murphy, D. M., & Koop, T. (2005). Review of the vapour pressures of ice and supercooled water for atmospheric applications. *Quarterly Journal of the Royal Meteorological Society*, *131*, 1539–1565.

Nelsen, R. B. (2007). *An introduction to copulas*. Berlin, Heidelberg: Springer-Verlag.

Ng, S. K., Krishnan, T., & McLachlan, G. J. (2012). *The EM algorithm*. Berlin, Heidelberg: Springer-Verlag.

Nimmegeers, P., Telen, D., Logist, F., & Van Impe, J. (2016). Dynamic optimization of biological networks under parametric uncertainty. *BMC systems biology*, *10*, 86.

Noh, Y., Choi, K., & Du, L. (2009). Reliability-based design optimization of problems with correlated input variables using a gaussian copula. *Structural and Multidisciplinary Optimization*, *38*, 1–16.

Oladyshkin, S., & Nowak, W. (2012). Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion. *Reliability Engineering & System Safety*, *106*, 179–190.

Pisano, R., Fissore, D., Barresi, A. A., Brayard, P., Chouvenc, P., & Woinet, B. (2013). Quality by design: optimization of a freeze-drying cycle via design space in case of heterogeneous drying behavior and influence of the freezing protocol. *Pharmaceutical Development and Technology*, *18*, 280–295.

Roffel, B., & Betlem, B. (2007). *Process dynamics and control: modeling for control and prediction*. Chichester (UK): John Wiley & Sons.

Rossner, N. (2014). *Robuste modellgestützte Prozessführung auf Basis von Gauß'schen Mischdichten am Beispiel der Bray-Liebhafsky-Reaktion und der autotrophen Kultivierung von Ralstonia eutropha H16*. Ph.D. thesis ,TU Berlin, Berlin.

Sahinidis, N. V. (2004). Optimization under uncertainty: state-of-the-art and opportunities. *Computers & Chemical Engineering*, *28*, 971–983.

Schenkendorf, R. (2014a). A general framework for uncertainty propagation based on point estimate methods. In *Second european conference of the prognostics and health management society, PHME14, Nantes, France*.

Schenkendorf, R. (2014b). *Optimal experimental design for paramter identification and model selection*. Ph.D. thesis ,Otto-von-Guericke-Universität, Magdeburg.

Schenkendorf, R., Xie, X., Rehbein, M., Scholl, S., & Krewer, U. (2018). The impact of global sensitivities and design measures in model-based optimal experimental design. *Processes*, *6*, 27.

Searles, J. (2004). Observation and implications of sonic water vapor flow during freeze-drying. *American Pharmaceutical Review*, *7*, 58–69.

Sheehan, P., & Liapis, A. I. (1998). Modeling of the primary and secondary drying stages of the freeze drying of pharmaceutical products in vials: Numerical results obtained from the solution of a dynamic and spatially multidimensional lyophilization model for different operational policies. *Biotechnology and Bioengineering*, *60*, 712–728.

Shi, J., Biegler, L. T., Hamdan, I., & Wassick, J. (2016). Optimization of grade transitions in polyethylene solution polymerization process under uncertainty. *Computers & Chemical Engineering*, *95*, 260–279.

Shi, X., Ye, M., Curtis, G. P., Miller, G. L., Meyer, P. D., Kohler, M., Yabusaki, S., & Wu, J. (2014). Assessment of parametric uncertainty for groundwater reactive transport modeling. *Water Resources Research*, *50*, 4416–4439.

Smith, R. C. (2013). *Uncertainty quantification: theory, implementation, and applications* volume 12. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics.

Srinivasan, B., Bonvin, D., Visser, E., & Palanki, S. (2003). Dynamic optimization of batch processes: Ii. role of measurements in handling uncertainty. *Computers & Chemical Engineering*, *27*, 27–44.

Tang, X. C., & Pikal, M. J. (2004). Design of freeze-drying processes for pharmaceuticals: practical advice. *Pharmaceutical Research*, *21*, 191–200.

Telen, D., Vallerio, M., Cabianca, L., Houska, B., Van Impe, J., & Logist, F. (2015). Approximate robust optimization of nonlinear systems under parametric uncertainty and process noise. *Journal of Process Control*, *33*, 140–154.

Vallerio, M., Hufkens, J., Van Impe, J., & Logist, F. (2015). An interactive decision-support system for multi-objective optimization of nonlinear

dynamic processes with uncertainty. *Expert Systems with Applications*, *42*, 7710–7731.

Vallerio, M., Telen, D., Cabianca, L., Manenti, F., Van Impe, J., & Logist, F. (2016). Robust multi-objective dynamic optimization of chemical processes using the sigma point method. *Chemical Engineering Science*, *140*, 201–216.

Wächter, A., & Biegler, L. T. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, *106*, 25–57.

Xie, X., Krewer, U., & Schenkendorf, R. (2018). Robust optimization of dynamical systems with correlated random variables using the point estimate method. *IFAC-PapersOnLine*, *51*, 427–432.

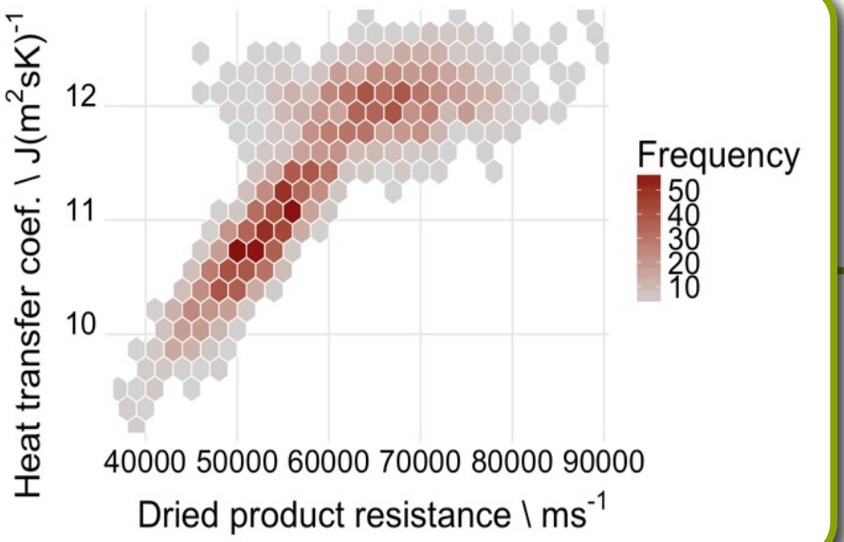
Xu, T., & Gómez-Hernández, J. J. (2016). Characterization of non-gaussian conductivities and porosities with hydraulic heads, solute concentrations, and water temperatures. *Water Resources Research*, *52*, 6111–6136.

Zhao, Y.-G., & Ono, T. (2001). Moment methods for structural reliability. *Structural Safety*, *23*, 47–75.

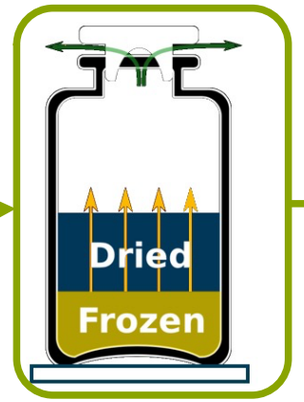
Robust Freeze-Drying Optimization

Problem

Uncertain model parameters

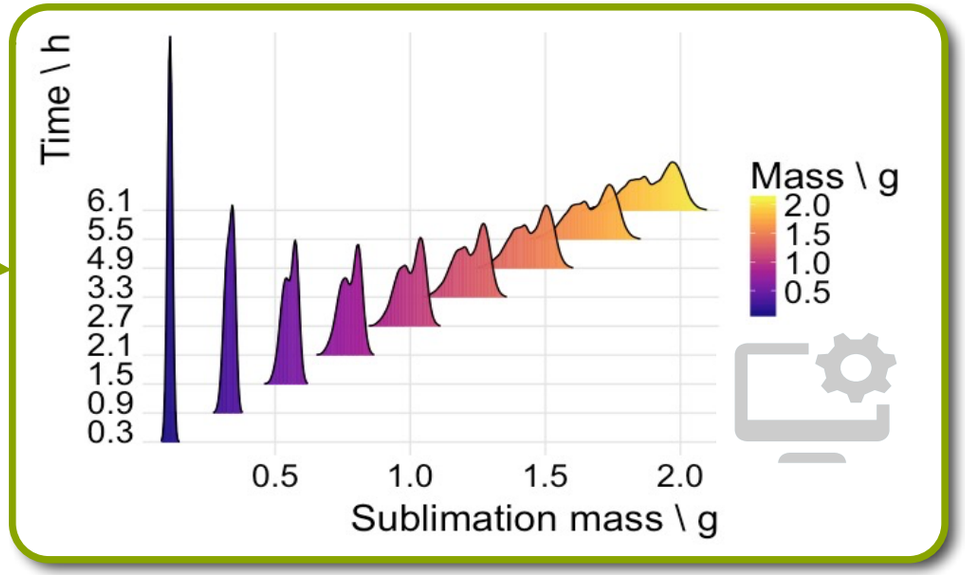


Freeze-drying process

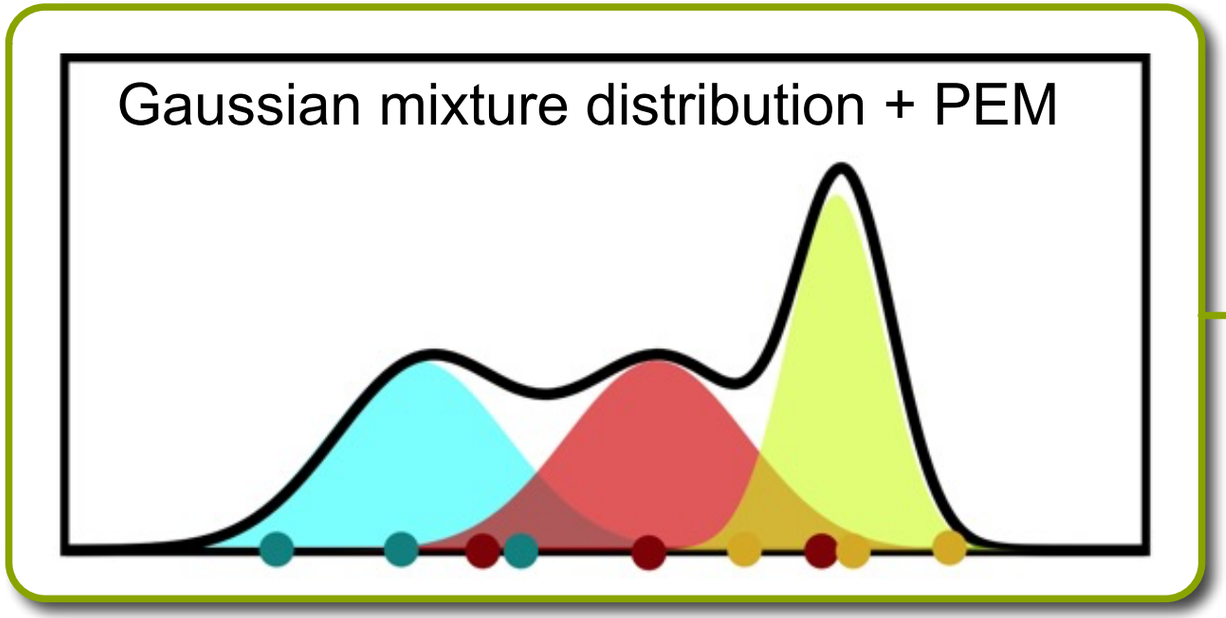


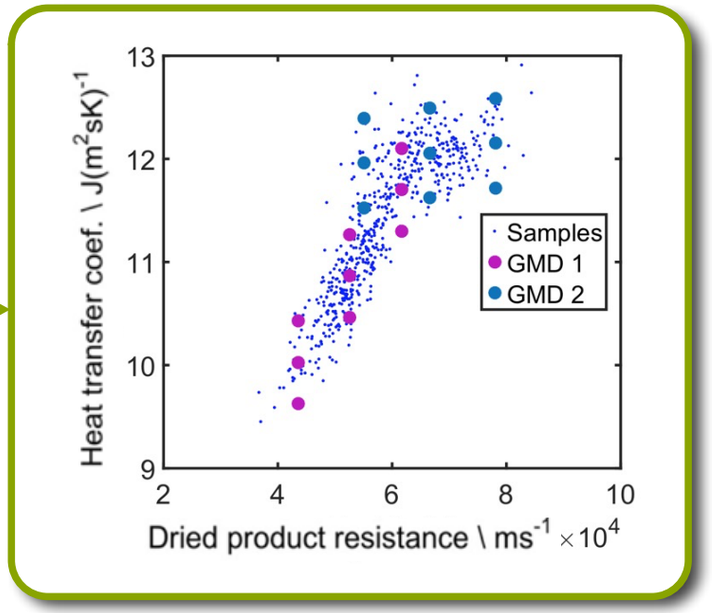
model

Uncertain process performance



Solution





Highlights:

* improved algorithm for robust process design is introduced to propagate and quantify complex parameter uncertainties and correlations efficiently

* a Gaussian mixture distribution based point estimated method is proposed to quantify non-Gaussian distributed uncertainties in the model parameters and model simulations

* robust process design concept has been successfully demonstrated with the freeze-drying process and compared critically with existing approaches