Bayesian Approach to Probabilistic Design Space Characterization: A Nested Sampling Strategy

Kennedy P. Kusumo,[†] Lucian Gomoescu,[†,‡] Radoslav Paulen,[¶] Salvador García Muñoz,[§] Constantinos C. Pantelides,[†,‡] Nilay Shah,[†] and Benoît Chachuat[*,†]

[†]*Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, London, SW7 2AZ, UK*

[‡]*Process Systems Enterprise Ltd, London, W6 7HA, UK*

[¶]*Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava, 812 43 Bratislava, Slovakia*

[§]*Small Molecule Design and Development, Lilly Research Laboratories, Eli Lilly & Company, Indianapolis, IN 46285, USA*

E-mail: b.chachuat@imperial.ac.uk

Abstract

Quality by design in pharmaceutical manufacturing hinges on computational methods and tools that are capable of accurate quantitative prediction of the design space. This paper investigates Bayesian approaches to design space characterization, which determine a feasibility probability that can be used as a measure of reliability and risk by the practitioner. An adaptation of nested sampling—a Monte Carlo technique introduced to compute Bayesian evidence—is presented. The nested sampling algorithm maintains a given set of live points through regions with increasing probability feasibility until reaching a desired reliability level. It furthermore leverages efficient strategies

1

from Bayesian statistics for generating replacement proposals during the search. Features and advantages of this algorithm are demonstrated by means of a simple numerical example and two industrial case studies. It is shown that nested sampling can outperform conventional Monte Carlo sampling and be competitive with flexibility-based optimization techniques in low-dimensional design space problems. Practical aspects of exploiting the sampled design space to reconstruct a feasibility probability map using machine learning techniques are also discussed and illustrated. Finally, the effectiveness of nested sampling is demonstrated on a higher-dimensional problem, in the presence of a complex dynamic model and significant model uncertainty.

Introduction

Over the years the pharmaceutical industry has identified an increasing need for systematic and holistic approaches to drug development and manufacturing, which has led to a high penetration of process systems engineering (PSE) tools.[1,2] This need is prompted by safety concerns and regulations alongside growing pressure to increase efficiency, both in production and in process development. A recent estimate[3] amounts to US$2.6Bn for a single novel drug put on the market. To improve practices in the industry the International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) introduced the Quality by Design (QbD) initiative through a series of five guidelines: ICH Q8–12.[4] Historically, pharmaceutical development and manufacturing had emphasized checklist-based operations rather than scientific understanding. So, by promoting a scientific and risk-based approach to pharmaceutical product development and manufacturing, the QbD initiative and ICH guidelines triggered a paradigm shift in the industry and a complete new range of activities for the practitioners.

The QbD approach defines a Quality Target Product Profile (QTPP), which is a prospective summary of the quality characteristics of the pharmaceutical product that ensures the desired quality, safety and efficacy. The (physical, chemical, biological or microbiological)

properties that should be within an appropriate limit, range or distribution to ensure the desired product quality are called Critical Quality Attributes (CQAs). The limits of acceptability of such CQAs is known as the Design Space (DS). More precisely, a DS consists of *"the multidimensional combination and interaction of input variables (material attributes) and process parameters that have been demonstrated to provide assurance of quality"*.[5] In practice a regulatory process of post-approval change is not required so long as the process parameters vary within the limits of the approved DS.

A classical approach to DS characterization entails the following four steps:[6,7] (i) conduct a thorough experimental investigation of the relationships between process parameters and CQAs; (ii) analyze the sensitivity of the process parameters on the CQAs to only retain those parameters presenting a significant sensitivity; (iii) establish a mathematical or graphical representation of the DS using data-driven modeling and optimization techniques; and (iv) test the final DS by running further validation experiments, prior to submitting it for approval to the regulatory agencies. Multivariate statistical techniques, e.g. based on latent-variable modeling, have been proposed to reduce the cost and time needed to conduct these experiments.[8,9] Nevertheless, the presence of a large number of external disturbances and/or potential control manipulations lead to high-dimensional DS, whose effective characterization can be empowered by the use of process models.[10] Such process models may be either data-driven or knowledge-driven insofar as they describe the relationships between process parameters and CQAs accurately.

Characterizing a DS based on a process model is akin to the popular concepts of resilience and flexibility introduced by the PSE community during the 1980s.[11,12] Two classical problems in this literature are: (i) the flexibility test,[13] which verifies that a feasible operation can be obtained for specific process parameters under a range of uncertainty scenarios—possibly by manipulating certain controls; and (ii) the flexibility index,[14] which represents the maximal deviation of the process parameters from given nominal values that can be tolerated in order for the operation to remain feasible under a range of uncertainty scenarios.

A related concept is that of process operability,[15,16] which starts from an estimated range of the process uncertainty and calculates the desired ranges of the control variables so that a control strategy can guarantee a feasible operation.

Like the result of any model-based computation, the reliability of any DS determined on the basis of a process model depends on the accuracy of that model. If the uncertain parameters of the model follow a certain probability distribution then the DS itself is not a well-delineated region but a probabilistic one, where each point in the process parameter space is characterized by a probability of meeting the QTPP.[17] Interestingly, a closely related concept to that of probabilistic DS was introduced in the PSE literature later during the 1990s under the name stochastic flexibility,[18,19] as a way of measuring the probability of a process design to operate feasibly in the presence of uncertainty. This probabilistic interpretation is in agreement with Peterson[20] who defines the DS as the set of process parameters such that the manufactured product satisfies the QTPP alongside other process constraints with a probability greater than a given threshold.

Existing computational approaches to probabilistic DS characterization differ in how they account for process model uncertainty and how they approximate the DS itself. Process model uncertainty is represented in either one of two ways: (i) set-membership, most commonly a joint confidence ellipsoid at a given confidence level as derived from a frequentist parameter estimation;[10,21] or (ii) sampled distribution, for instance the sampled parameter posterior in Bayesian estimation or a bootstrap parameter distribution.[9,20,22] Caution must be exerted in both cases since assuming an ellipsoidal region for non-normally distributed parameters or representing a probability distribution with an insufficient number of samples can lead to large errors in the predicted DS.

In terms of DS representation, one may broadly categorize the existing approaches as: (i) design-centering algorithms; and (ii) sampling algorithms. The former consider a class of parameterized shapes—most commonly a box—and transform the DS characterization into an optimization problem that seeks to inscribe the largest possible shape within the DS.

The flexibility-index problem[14,23] falls into this category and it was recently applied to probabilistic DS computation.[21,24] A drawback with this approach is that a simple shape may not provide a good approximation of the DS and thus introduce significant conservatism. Another drawback is that design-centering problems give rise to complex mathematical programs with either robust (semi-infinite) or chance constraints that are computationally hard to tackle rigorously.[25,26] By contrast, sampling algorithms discretize the process parameter range and return a subset of the samples that satisfy the QTPP and other constraints up to the desired reliability value. Exhaustive sampling may be achieved via a fine uniform gridding or (quasi) Monte Carlo sampling. The assessment of each parameter sample can be made via the solution of an optimization problem, which is akin to a flexibility test and is especially suited to a set-membership description of the model uncertainty.[17,21,27] If the model parameter uncertainty is represented by a sampled distribution instead, both Monte Carlo and Bayesian techniques can be used for propagating the uncertainty to the CQAs and estimate a feasibility probability.[9,22] These sampling techniques have proven effective in practice but they are computationally expensive and mainly tractable for low-dimensional DS at present. For computationally demanding process models surrogate-based approaches can be applied, including kriging, radial basis functions, and high-dimensional model representation.[28] For instance, surrogate-based adaptive sampling[29–31] has been shown to speed-up the computation of flexibility indices. But while it can be computationally cheaper to characterize the DS using a simple surrogate, one needs to account for the additional burden of constructing the surrogate itself as well as any additional approximation error that may be introduced.

Despite the large body of research on DS computation, there is a clear need for algorithms with improved computational efficiency to tackle industrially-relevant problems that have more design parameters or greater uncertainty. The focus of this paper is on nested sampling, a Monte Carlo technique that was introduced by Skilling[32] for estimating the Bayesian evidence in parameter estimation. This algorithm proceeds by progressively sampling in

nested contours of increasing likelihood, so as to maintain a dense enough sample in regions of higher likelihood, in the manner of adaptive sampling. An advantage of nested sampling over Markov chain Monte Carlo (MCMC) techniques is its better ability to handle multimodal posteriors.[33]

Our main contribution herein is an adaptation of nested sampling for the characterization of a probabilistic DS. In this context nested sampling determines a set of samples with a feasibility probability larger than a given reliability threshold. A by-product of the algorithm is a second set of samples with a feasibility probability below the desired threshold, which provides an approximation of the entire probabilistic DS. In this manner nested sampling can offer a greater flexibility in selecting the probability threshold to provide assurance of quality. An important motivation behind the adaptation of nested sampling is the availability of efficient strategies for generating replacement proposals as the algorithm progresses,[33,34] which we can leverage to the benefit of DS characterization. Another appeal is the ability to readily exploit a sampled joint posterior distribution of the model parameters: by combining the proposed approach with a Bayesian estimation procedure for the model parameters, one can arrive at a truly Bayesian approach to DS characterization directly from experimental data. Last but not least importantly nested sampling is, like other sampling-based techniques, non-intrusive in the sense that it relies on the result of model simulations at given process parameter values only. In principle, this allows for black-box models such as a process flowsheet or a CFD model to be used for DS characterization.

The rest of the paper is organized as follows: In the following section we review the mathematical formulation of the design space characterization problem. Then we present the nested sampling approach and use a simple example to illustrate its main features. We also test the method on two case studies of increasing complexity and discuss the results, before concluding the paper.

Problem Statement

Consider a manufacturing process for a pharmaceutical product that has its quality defined by some CQAs, denoted by $\boldsymbol{s} \in \mathbb{R}^{n_s}$. Assume that a mathematical model of the process (either knowledge- or data-driven) is available that predicts the CQAs corresponding to a set of process parameters, denoted by $\boldsymbol{d} \in \mathcal{K}$ within the *knowledge space* $\mathcal{K} \subset \mathbb{R}^{n_d}$:

$$\boldsymbol{s} = \boldsymbol{f}\left(\boldsymbol{d}, \boldsymbol{\theta}\right) \tag{1}$$

The model parameters, $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$ may represent physical constants, coefficients in a regression model, or disturbances that affect the CQAs. The mapping $\boldsymbol{f} : \mathbb{R}^{n_d} \times \mathbb{R}^{n_\theta} \to \mathbb{R}^{n_s}$ needs not be given in closed-form but could be implicitly defined by a set of algebraic or differential equations or even a black-box functions such as a process simulator. Assume furthermore that the constraints on the CQAs are represented alongside other process constraints by the following inequalities:

$$\boldsymbol{G}\left(\boldsymbol{d}, \boldsymbol{\theta}\right) := \boldsymbol{g}\left(\boldsymbol{d}, \boldsymbol{f}\left(\boldsymbol{d}, \boldsymbol{\theta}\right)\right) \leq \boldsymbol{0} \tag{2}$$

Notice that the variable $\boldsymbol{s}$ can be abstracted away from the inequality constraints without loss of generality. Similarly to the mapping $\boldsymbol{f}$ above, the constraint function $\boldsymbol{G} : \mathbb{R}^{n_d} \times \mathbb{R}^{n_\theta} \to \mathbb{R}^{n_g}$ is not necessarily available in closed form.

Ignoring the uncertainty in the model parameters leads to a nominal design space:

$$\mathcal{D}_{\mathrm{nom}} := \{\boldsymbol{d} \in \mathcal{K} : \boldsymbol{G}(\boldsymbol{d}, \boldsymbol{\theta}_{\mathrm{nom}}) \leq \boldsymbol{0}\} \tag{3}$$

for a given nominal value $\boldsymbol{\theta}_{\mathrm{nom}}$ of the model parameters. However, the value of $\boldsymbol{\theta}$ is inherently uncertain in practice by the nature of the modeling exercise. A Bayesian framework considers $\boldsymbol{\theta}$ as random variables with a joint distribution $p(\boldsymbol{\theta})$ that describes the belief on the value of $\boldsymbol{\theta}$. For instance, $p(\boldsymbol{\theta})$ could be estimated from experimental data using Bayesian inference. In

7

this framework the model may only be used to predict the probability that the manufacturing process is feasible for a given $\boldsymbol{d} \in \mathcal{K}$, which is akin to a stochastic flexibility test:[18]

$$\mathbb{P}\left[\boldsymbol{G}(\boldsymbol{d}, \cdot) \leq \boldsymbol{0} \mid p(\boldsymbol{\theta})\right] := \int_{\{\boldsymbol{\theta}: \boldsymbol{G}(\boldsymbol{d}, \boldsymbol{\theta}) \leq \boldsymbol{0}\}} p(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta} \tag{4}$$

The problem of interest throughout this paper is to determine the probabilistic DS given by:

$$\mathcal{D}_\alpha := \{\boldsymbol{d} \in \mathcal{K} : \mathbb{P}\left[\boldsymbol{G}(\boldsymbol{d}, \cdot) \leq \boldsymbol{0} \mid p(\boldsymbol{\theta})\right] \geq \alpha\} \tag{5}$$

where $0 < \alpha \leq 1$ is the so-called reliability value.[20] Notice that the following set-membership counterpart to the probabilistic DS is often computed instead of $\mathcal{D}_\alpha$ in practice:[21]

$$\widehat{\mathcal{D}}_\alpha := \{\boldsymbol{d} \in \mathcal{K} : \forall \boldsymbol{\theta} \in \Theta_\alpha, G(\boldsymbol{d}, \boldsymbol{\theta}) \leq \boldsymbol{0}\} \tag{6}$$

with Θ_α chosen as the highest posterior density (HPD) set such that $\int_{\boldsymbol{\theta} \in \Theta_\alpha} p(\boldsymbol{\theta}) = \alpha$. However, $\mathcal{D}_\alpha$ and $\widehat{\mathcal{D}}_\alpha$ are not equivalent in general when the constraints are nonlinear or the model parameters are not normally distributed. ~~It is also worth noting an important assumption underlying the definition of $\mathcal{D}_\alpha$ (or $\widehat{\mathcal{D}}_\alpha$), which is that the process model is either structurally correct or that any structural mismatch can be reported in terms of parametric uncertainty~~

According to the monotonicity property that $\mathcal{D}_\alpha \supset \mathcal{D}_{\alpha'}$ whenever $\alpha < \alpha'$, a higher reliability value increases conservatism by shrinking the DS. A practical choice for α entails trading-off the risk of violating QTPP against the loss of operational flexibility: a lower α value increases false positives in $\mathcal{D}_\alpha$ by including design parameters that will not fulfill all of the QTPP and process constraints as expected; while a higher α value increases false negatives by excluding design parameters that are in fact feasible. Because false positives pose a threat to the assurance of quality, practitioners are prompted to be conservative. But the choice of α remains specific to the nature of the process at hand and to the risks involved.

It is also worth noting an important assumption underlying the definition of $\mathcal{D}_\alpha$ (or $\widehat{\mathcal{D}}_\alpha$),

namely that the process model is either structurally correct or that any structural mismatch can be reported in terms of parametric uncertainty. A practitioner's lacking confidence in the process model structure can always increase conservatism by opting for a higher reliability value α. However, the probabilities conveyed by $\mathcal{D}_\alpha$ could become misleading in the presence of large structural mismatch. A possible remedy entails the consideration of multiple candidate models under the Bayesian framework and the restatement of $\mathbb{P}\left[\boldsymbol{G}(\boldsymbol{d}, \cdot) \leq \boldsymbol{0} \,|\, p(\boldsymbol{\theta})\right]$ to explicitly consider structural uncertainty.

Illustrative Example

We consider a simple case study with two design variables, $\boldsymbol{d} := (d_1, d_2)^\intercal$ and a single CQA, s. We assume the following relationship between the design variables and the CQA:

$$s = \theta d_1^2 + d_2 \tag{7}$$

with the model parameter θ. The goal is to characterize the probabilistic DS inside the knowledge space $\mathcal{K} := [-1, 1]^2$ imposed by the following QTPP:

$$0.20 \leq s \leq 0.75 \tag{8}$$

We consider the nominal case to be given by $\theta_{\mathrm{nom}} = 1$, which gives the following nominal design space:

$$\mathcal{D}_{\mathrm{nom}} := \{\boldsymbol{d} \in [-1, 1]^2 : 0.20 \leq d_1^2 + d_2 \leq 0.75\} \tag{9}$$

Under a normality assumption for the model parameters, $\theta \sim \mathcal{N}(\mu_\theta, \sigma_\theta)$ the probabilistic DS
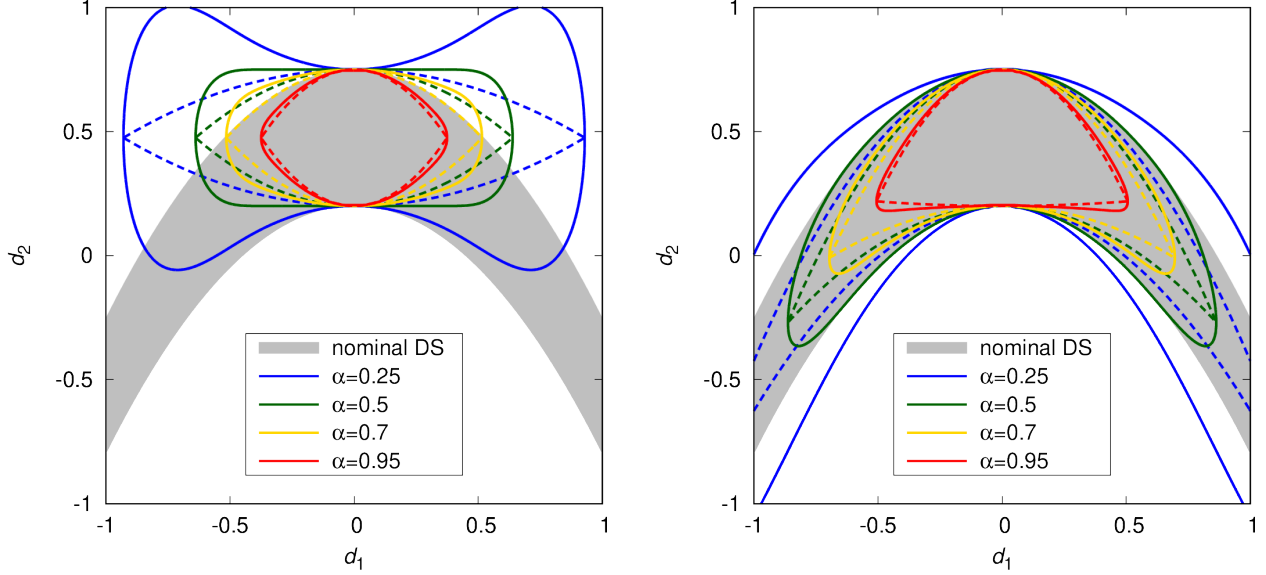
Figure 1: Depiction of the design space for the illustrative example. The nominal design space in Equation (9) is represented with the gray-shaded area. The probabilistic design spaces corresponding to different reliability values α in Equation (10) are shown in solid lines with different colors. The set-membership counterparts in Equation (11) are shown in dashed lines. The left and right plots are for the uncertainty scenarios $\theta \sim \mathcal{N}(0,1)$ and $\theta \sim \mathcal{N}(1, \sqrt{0.3})$, respectively.

can be expressed analytically as:

$$\mathcal{D}_\alpha := \left\{ \boldsymbol{d} \in [-1,1] : \mathbb{P}\left[0.20 \leq \theta d_1^2 + d_2 \leq 0.75 \,\middle|\, \theta \sim \mathcal{N}(\mu_\theta, \sigma_\theta) \right] \geq \alpha \right\} \tag{10}$$

$$= \left\{ \boldsymbol{d} \in [-1,1] : \mathrm{erf}\left(\frac{0.75 - \mu_\theta d_1^2 - d_2}{\sqrt{2}\sigma_\theta d_1^2} \right) - \mathrm{erf}\left(\frac{0.20 - \mu_\theta d_1^2 - d_2}{\sqrt{2}\sigma_\theta d_1^2} \right) \geq 2\alpha \right\}$$

Figure 1 compares $\mathcal{D}_\alpha$ for two parameter uncertainty scenarios, $\theta \sim \mathcal{N}(0,1)$ in the left plot and $\theta \sim \mathcal{N}(1, \sqrt{0.3})$ in the right plot, and different reliability values α. As expected, a more precise and accurate description of the model uncertainty enables a more favorable trade-off QTPP satisfaction and operational flexibility. With an imprecise and inaccurate uncertainty description (left plot), avoiding false positives requires a high reliability value (e.g. $\alpha = 0.95$), but this results in a small DS and many false negatives. As the precision and accuracy of the model parameter improves (right plot), a lower reliability value may be selected without generating false positives (e.g. $\alpha = 0.5$) and at the same time enable a

much larger DS.

For comparison, Figure 1 also shows the DS counterpart for a set-membership uncertainty description based on Equation (6), which can also be expressed analytically as:

$$\widehat{\mathcal{D}}_\alpha := \left\{ \boldsymbol{d} \in [-1, 1] : \forall \theta \in \mu_\theta \pm z_\alpha \sigma_\theta, 0.20 \leq \theta d_1^2 + d_2 \leq 0.75 \right\} \tag{11}$$

$$= \left\{ \boldsymbol{d} \in [-1, 1] : 0.20 - (\mu_\theta - z_\alpha \sigma_\theta) d_1^2 \leq d_2 \leq 0.75 - (\mu_\theta + z_\alpha \sigma_\theta) d_1^2 \right\}$$

with the quantile $z_\alpha = \sqrt{2}\,\mathrm{erf}^{-1}(\alpha)$. For the same reliability value α, $\widehat{\mathcal{D}}_\alpha$ is generally more conservative than $\mathcal{D}_\alpha$. And the mismatch between $\mathcal{D}_\alpha$ and $\widehat{\mathcal{D}}_\alpha$ furthermore reduces as α gets closer to 1. The rest of the paper focuses on computational approach to characterizing the Bayesian design space $\mathcal{D}_\alpha$.

Methodology

Our main focus is on computational methods that rely on sampling to characterize a probabilistic design space. These methods are appealing in that they allow for the use of a process model in explicit or implicit form and they can accommodate any probability distribution for the model parameters. A pseudo-code for the standard Monte Carlo approach is presented in Algorithm 1, where both the process parameter space range and the model uncertainty distribution are discretized. At each process parameter sample, $\boldsymbol{d}_i \in \mathcal{S}_d$ the probability of QTTP satisfaction is estimated via an ensemble of simulations over the uncertainty scenarios in $\mathcal{S}_\theta$:

$$\widehat{\mathbb{P}}[\boldsymbol{G}(\boldsymbol{d}_i, \cdot) \leq \boldsymbol{0} \mid \mathcal{S}_\theta] := \sum_{(\boldsymbol{\theta}_j, \omega_j) \in \mathcal{S}_\theta} \mathbb{1}[\boldsymbol{G}(\boldsymbol{d}_i, \boldsymbol{\theta}_j)]\,\omega_j \tag{12}$$

where $\mathbb{1}[\cdot]$ stands for the indicator function, such that $\mathbb{1}[\boldsymbol{g}] = 1$ if $g_k \underset{\sim}{\leq} 0, \forall k$ and $\mathbb{1}[\boldsymbol{g}] = 0$ otherwise; and the samples $\boldsymbol{\theta}_j$ with corresponding weights ω_j are drawn from the uncertainty distribution $p(\cdot)$. Although effective this approach can prove computationally prohibitive be-

11

Algorithm 1 Standard Monte Carlo Sampling for Design Space Characterization
---
1: **input** $\mathcal{S}_d = \{\boldsymbol{d}_i \in \mathcal{K} : i = 1, \ldots, N_d\}$, $\mathcal{S}_\theta = \{(\boldsymbol{\theta}_j, \omega_j) \sim p(\cdot) : j = 1, \ldots, N_\theta\}$
2: $\mathcal{DS} \leftarrow \emptyset$
3: **for all** $\boldsymbol{d}_i \in \mathcal{S}_d$ **do**
4: $\quad \mathcal{DS} \leftarrow \mathcal{DS} \cup \left\{ \left(\boldsymbol{d}_i \,;\, \widehat{\mathbb{P}}[\boldsymbol{G}(\boldsymbol{d}_i, \cdot) \le \boldsymbol{0} \mid \mathcal{S}_\theta] \right) \right\}$
5: **end for**
6: **return** $\mathcal{DS}$
---

cause the number of model simulations grows as the product between the numbers of process parameter samples, N_d and uncertainty scenarios, N_θ. A major challenge for making such methods more efficient thus entails reducing the number of model simulations. The nested sampling algorithm described next aims precisely at sampling the feasibility probability map more uniformly in order to generate a larger number of points with high reliability value.

Nested Sampling for Design Space Characterization

Nested sampling[32] is a Monte Carlo technique for estimating the evidence in Bayesian parameter estimation. Specifically, given a prior distribution $\pi(\boldsymbol{\theta})$ and a likelihood function $\mathcal{L}(\boldsymbol{\theta})$, the posterior distribution $p(\boldsymbol{\theta})$ can be inferred from Bayes' theorem:

$$p(\boldsymbol{\theta}) = \frac{\mathcal{L}(\boldsymbol{\theta}) \, \pi(\boldsymbol{\theta})}{Z}$$

where $Z := \mathbb{E}[\mathcal{L}(\boldsymbol{\theta}) \mid \boldsymbol{\theta} \sim \pi(\cdot)] = \int \cdots \int_{\mathbb{R}^{n_\theta}} \mathcal{L}(\boldsymbol{\theta}) \, \pi(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}$ denotes the Bayesian evidence. The algorithm proceeds by progressively sampling in nested contours of increasing likelihood, so as to maintain a dense enough sample in regions of higher likelihood—refer to Appendix A of the Supplementary Information~~the appendix~~ for further details. This feature is an important motivation for investigating nested sampling in probabilistic DS applications. Before delving into the details of the algorithm, a quick comparison of nested sampling for parameter estimation and DS characterization is presented in Table 1.

A pseudo-code of nested sampling for probabilistic DS characterization is presented in Algorithm 2. The inputs to this algorithm comprise (line 1): the knowledge space, $\mathcal{K}$; the

Table 1: Comparison of nested sampling for parameter estimation and design space characterization.

Bayesian parameter estimation	Bayesian design space characterization
Outcome	
An estimate of the Bayesian evidence Z	A set of samples of given size within the probabilistic design space at the desired reliability value α^*, $\{\boldsymbol{d}_i \in \mathcal{D}_{\alpha^*} : i = 1, \ldots, N_d\}$
A set of weighted samples from the posterior, $\{(\boldsymbol{\theta}_j, \omega_j) \sim p(\cdot) : j = 1, \ldots, N_\theta\}$	A set of samples at a lesser reliability value $\alpha < \alpha^*$, $\{\boldsymbol{d}_k \in \mathcal{D}_{\alpha_k} : \alpha_k < \alpha^*, k = 1, \ldots, N'_d\}$
Challenge	
How to efficiently sample the prior distribution subject to a likelihood constraint, $\mathcal{L}(\boldsymbol{\theta}) > \mathcal{L}_{\min}$ so as to have a high acceptance rate and uniform spread	How to efficiently sample the design space subject to a feasibility probability constraint, $\widehat{\mathbb{P}}[\mathbf{G}(\boldsymbol{d}, \cdot) \leq 0] > \widehat{\mathbb{P}}_{\min}$ so as to have a high acceptance rate and uniform spread

desired reliability value α^*; the number of replacement candidates per iteration N_{R}; an initial sample set of design parameters of size N_{L} within the knowledge space $\mathcal{S}_{\mathrm{L}} = \{\boldsymbol{d}_i \in \mathcal{K} : i = 1, \ldots, N_{\mathrm{L}}\}$; and set of weighted samples of size N_θ describing the model parameter uncertainty $\mathcal{S}_\theta = \{(\boldsymbol{\theta}_j, \omega_j) \sim p(\cdot) : j = 1, \ldots, N_\theta\}$. Recall that the latter may either be drawn from a closed-form distribution (with equal weights), e.g. a Gaussian distribution with a given mean and covariance matrix, or be the result of a Bayesian parameter estimation using MCMC or nested sampling again.

The N_{L} points in the set $\mathcal{S}_{\mathrm{L}}$ are called *live points* in the nested sampling literature and may be initialized via uniform sampling in the knowledge space $\mathcal{K}$. Nested sampling starts by estimating the feasibility probability at each of these points according to Equation (12), which uses the discretized uncertainty set $\mathcal{S}_\theta$. Then each iteration of the algorithm proceeds by sampling $N_{\mathrm{R}} \geq 1$ new points within an envelope that encloses the current live points and substituting them with the live point having the least feasibility probability in case of improvement. The replaced live points are called *dead points* and stored with their feasible probability in the result set $\mathcal{DS}$. Notice that the main benefit of using $N_{\mathrm{R}} > 1$ is in regards of vectorizing the feasibility probability evaluation, although this parallelization capability

Algorithm 2 Nested Sampling for Design Space Characterization

---

1: **input** $\mathcal{K}$, $\mathcal{S}_{\mathrm{L}} = \{\boldsymbol{d}_i \in \mathcal{K} : i = 1, \ldots, N_{\mathrm{L}}\}$, $\mathcal{S}_\theta = \{(\boldsymbol{\theta}_j, \omega_j) \sim p(\cdot) : j = 1, \ldots, N_\theta\}$, α^*, N_{R}
2: $\mathcal{DS} \leftarrow \emptyset$
3: **while** $\exists \boldsymbol{d}_i \in \mathcal{S}_{\mathrm{L}} : \widehat{\mathbb{P}}[\boldsymbol{G}(\boldsymbol{d}_i, \cdot) \leq \boldsymbol{0} \mid \mathcal{S}_\theta] < \alpha^*$ **do**
4:     $\widehat{\mathbb{P}}_{\min} \leftarrow \min\{\widehat{\mathbb{P}}[\boldsymbol{G}(\boldsymbol{d}_i, \cdot) \leq \boldsymbol{0} \mid \mathcal{S}_\theta] : d_i \in \mathcal{S}_{\mathrm{L}}\}$
5:     $\boldsymbol{d}_{\min} \leftarrow \arg\min\{\widehat{\mathbb{P}}[\boldsymbol{G}(\boldsymbol{d}_i, \cdot) \leq \boldsymbol{0} \mid \mathcal{S}_\theta] : d_i \in \mathcal{S}_{\mathrm{L}}\}$
6:     $\mathcal{S}_{\mathrm{R}} \leftarrow \{\boldsymbol{d}_k \in \mathcal{K} : k = 1, \ldots, N_{\mathrm{R}}\}$
7:     **for all** $\boldsymbol{d}_k \in \mathcal{S}_{\mathrm{R}}$ **do**
8:         **if** $\widehat{\mathbb{P}}[\boldsymbol{G}(\boldsymbol{d}_k, \cdot) \leq \boldsymbol{0} \mid \mathcal{S}_\theta] > \widehat{\mathbb{P}}_{\min}$ **then**
9:             $\mathcal{S}_{\mathrm{L}} \leftarrow \mathcal{S}_{\mathrm{L}} \cup \{\boldsymbol{d}_k\} \setminus \{\boldsymbol{d}_{\min}\}$
10:            $\mathcal{DS} \leftarrow \mathcal{DS} \cup \left\{\left(\boldsymbol{d}_{\min} ; \widehat{\mathbb{P}}_{\min}\right)\right\}$
11:         **end if**
12:     **end for**
13: **end while**
14: **for all** $\boldsymbol{d}_i \in \mathcal{S}_{\mathrm{L}}$ **do**
15:     $\mathcal{DS} \leftarrow \mathcal{DS} \cup \left\{\left(\boldsymbol{d}_i ; \widehat{\mathbb{P}}[\boldsymbol{G}(\boldsymbol{d}_i, \cdot) \leq \boldsymbol{0} \mid \mathcal{S}_\theta]\right)\right\}$
16: **end for**
17: **return** $\mathcal{DS}$

---

is not used for the case studies in this paper.

Key to the efficiency of nested sampling is the ability to generate replacement candidates that have a satisfactory acceptance rate (line 6). An important motivation behind the adaptation of nested sampling for DS characterization is that a variety of techniques have been developed by the Bayesian estimation community over the years to generate such points. This includes: (i) sampling from an enlarged ellipsoid or multiple enlarged ellipsoids enclosing the current live points;[33,34] or (ii) running a short MCMC from a randomly selected live point.[33,35]

Estimating the feasibility probability of the constraints for given process parameter values requires N_θ process model runs and is by far the most computationally demanding aspect of the algorithm. For N_{iter} nested sampling iterations, the total number of process model runs amounts to $N_\theta(N_{\mathrm{L}} + N_{\mathrm{iter}} N_{\mathrm{R}})$. An acceleration strategy consists in interrupting the evaluation of the feasibility probability $\widehat{\mathbb{P}}[\boldsymbol{G}(\boldsymbol{d}_k, \cdot) \leq \boldsymbol{0} \mid \mathcal{S}_\theta]$ of a replacement candidate $\boldsymbol{d}_k$ (line 8) if the feasibility probability mass accumulated so far and the mass of the remaining samples add up to less than $\widehat{\mathbb{P}}_{\min}$ (line 4)—e.g. by summing in decreasing order of probability weights ω_j

in Equation (12) for maximal efficiency. Such an interruption is indeed a guarantee that the replacement candidate should be rejected. Of course, the consequence is that the feasibility probability $\widehat{\mathbb{P}}[\boldsymbol{G}(\boldsymbol{d}_k, \cdot) \leq \boldsymbol{0} \mid \mathcal{S}_\theta]$ of any rejected point will not be estimated accurately when this acceleration strategy is used. This is the main reason why we do not append the rejected points to the result set $\mathcal{DS}$ in Algorithm 2, which is in agreement with our primary objective of uncovering N_L points with a feasibility probability greater than α^*.

The main iteration terminates when the feasibility probability of all the live points is no less than the target reliability value α^*. All of these points are appended with their corresponding feasibility probabilities to the result set $\mathcal{DS}$ (line 15) that already contains the dead points. It may happen that the number of points with a feasibility probability greater than α^* in $\mathcal{DS}$ is larger than N_L when multiple live point replacements occur during the final iteration. Also, Algorithm 2 will terminate prematurely in the event that $\mathcal{D}_{\alpha^*}$ is empty, which occurs when the model uncertainty is too large or the reliability value α^* is too high. This situation can nonetheless be detected, e.g. by monitoring the rate of improvement in the feasibility probability over a given number of iterations, and the algorithm can be stopped when progress is too slow.

An implementation of Algorithm 2 in a Python package called DEUS (standing for *DEsign under Uncertainty using Sampling techniques*) is available as part of the Supplementary Material to the paper. At each iteration the replacement candidates are generated by sampling in a single ellipsoid enclosing the current live points. Accordingly, the tuning parameters for a probabilistic DS computation in DEUS include: (i) the number of live points, N_L; (ii) the number of replacement candidates at each iteration, N_R; (iii) the initial enlargement factor of the ellipsoid; and (iv) the shrinking rate of that enlargement factor at each iteration. This package is used to solve all of the numerical case studies presented below, using as default parameters an initial enlargement factor of 30% and a shrinking rate of 0.2 for the ellipsoids[33] as well as $N_\mathrm{R} = 8$ replacement proposals. Despite the fact that no parallelization is implemented in DEUS at present, choosing an N_R value greater than 1 can reduce the over-

head caused by reconstructing an ellipsoid around the live points too frequently. Finally, an appropriate N_L value is highly dependent on the volume ratio between target design space and knowledge space, with a larger volume ratio calling for a larger N_L to maintain a suitable sample density across the DS. In our experience it is advisable to start with a low N_L to establish that $\mathcal{D}_{\alpha^*}$ is not empty then refine the characterization by increasing N_L. Like any other sampling-based approach, there is no guarantee that DEUS will not miss part of the feasible region. But the likelihood of this happening can be reduced by either increasing the enlargement factor of the ellipsoids, decreasing the related shrinking rate, or increasing the number of replacement proposals per iteration.

Exploitation of the Results

Sampling-based approaches to probabilistic DS characterization return a set of points and their corresponding feasibility probabilities. Turning these results into a format that can be exploited by the practitioner is clearly important. It is straightforward to represent a probabilistic design space in one or two dimensions graphically, for instance using a range of colors to depict different feasibility probability levels. This representation may also be used in three or four dimensions by splitting the range of the extra dimensions onto a trellis chart—see for instance the Suzuki reaction case study below.

But, more generally, a graphical presentation might be insufficient for the practitioner to easily test the feasibility probability of a given set of design parameters or to convey the resulting design space to third-parties in a concise way. Alternative representations include:

- Inscribing a simple shape—e.g. a box or an ellipsoid—within the envelope defined by the samples above a given reliability level in the data set $\mathcal{DS}$ returned by Algorithm 1 or 2. Under the assumption that the samples describe a convex set, one may proceed by first constructing a polyhedral representation of the set then fitting the desired shape within that polyhedron. The latter problem may be solved efficiently using convex optimization, but constructing a set of hyperplanes to describe the convex hull

of a set of points can be computationally burdensome (NP-hard).[36] This approach is furthermore inadequate for feasibility probability maps presenting nonconvex iso-contours or multiple modes.

- Fitting a non-parametric model to approximate the full feasibility probability map across the knowledge space. For instance, a a multilayer perceptron[37,38] (MLP)—a class of feed-forward artificial neural network with multiple (hidden) layers and non-linear activation functions—can be trained on the labelled data set $\mathcal{DS}$. The trained MLP provides a computationally cheap surrogate, transforming the DS samples into a suitable form for exploitation. It may be used readily to predict the feasibility probability for any design parameters $\boldsymbol{d} \in \mathcal{K}$. It may also be embedded into a design-centering optimization problem[26] for finding a subset of points, e.g. in the form of a box or an ellipsoid, with feasibility probability above a desired reliability value. This is supported by recent advances in deterministic global optimization with neural networks embedded.[39]

- Fitting a multinomial classifier[37] to separate the knowledge space into two or more subregions corresponding to different feasibility probability ranges—e.g. below and above the reliability value α^*. Similar to non-parametric regression an MLP has the ability to distinguish data sets that are not linearly separable and may be trained on the labelled data set $\mathcal{DS}$. The trained classifier then provides a cheap way of estimating the probability range of any design parameters $\boldsymbol{d} \in \mathcal{K}$ and a softmax function may be used in the final layer of the MLP to estimate the probability of $\boldsymbol{d} \in \mathcal{K}$ belonging to a given range. Like a non-parametric regression model such a classifier may also be embedded into a design-centering problem.

Our focus in the remainder of the paper is on MLP to approximate the full feasibility probability map. The MLP of interest has a single neuron in its output layer—whose state represents the feasibility probability—and its input layer comprises exactly n_d neurons—

one for each process parameter. A key, yet arduous, decision is selecting the numbers of hidden layers and hidden neurons in the MLP. At least one hidden layer is necessary since the feasibility probability map is generally nonlinear. An MLP with a single hidden layer is capable of approximating any continuous functions under mild assumptions on the activation function but it can take an arbitrary large number of neurons in that hidden layer to meet a desired accuracy.[40] Deep learning mitigates this problem by including additional hidden layers. However, it is important to keep in mind that the feasibility probability is inherently noisy due to the model uncertainty discretization. Therefore, the MLP may also be called upon to play a role in filtering this noise and avoiding overfitting of the data by restricting its number of hidden layers or neurons. Since the numerical case studies below have a handful of process parameters we consider hidden layers with a few dozen neurons as a rule of thumb. A more systematic analysis is beyond the scope of this paper.

Illustrative Example (Continued)

We revisit our simple case study concerned with the determination of the design space defined by the CQA constraint (8) using the uncertain model (7). The left plots in Figure 2 are the results of nested sampling (Algorithm 2) with a reliability target of $\alpha^* = 0.95$, $N_{\mathrm{L}} = 500$ live points, and $N_\theta = 100$ samples drawn from either $\mathcal{N}(0,1)$ (top plot) or $\mathcal{N}(1, \sqrt{0.3})$ (bottom plot). Overall, the resulting samples—depicted in a different color according to reliability membership—are in excellent agreement with the exact subregions delimited by the solid contours (Equation 10). Nevertheless a few of the points are misclassified due to the approximation of the feasibility probability based on Equation (12) which uses a finite number of uncertainty samples. This discrepancy can be removed by increasing the number of uncertainty samples N_θ, yet at the cost of a higher computational burden since the total number of model evaluations is directly proportional to N_θ.

The right plots in Figure 2 are generated using the Monte Carlo Algorithm 1 with Sobol sampling. These results are again in excellent agreement with the exact subregions, despite a
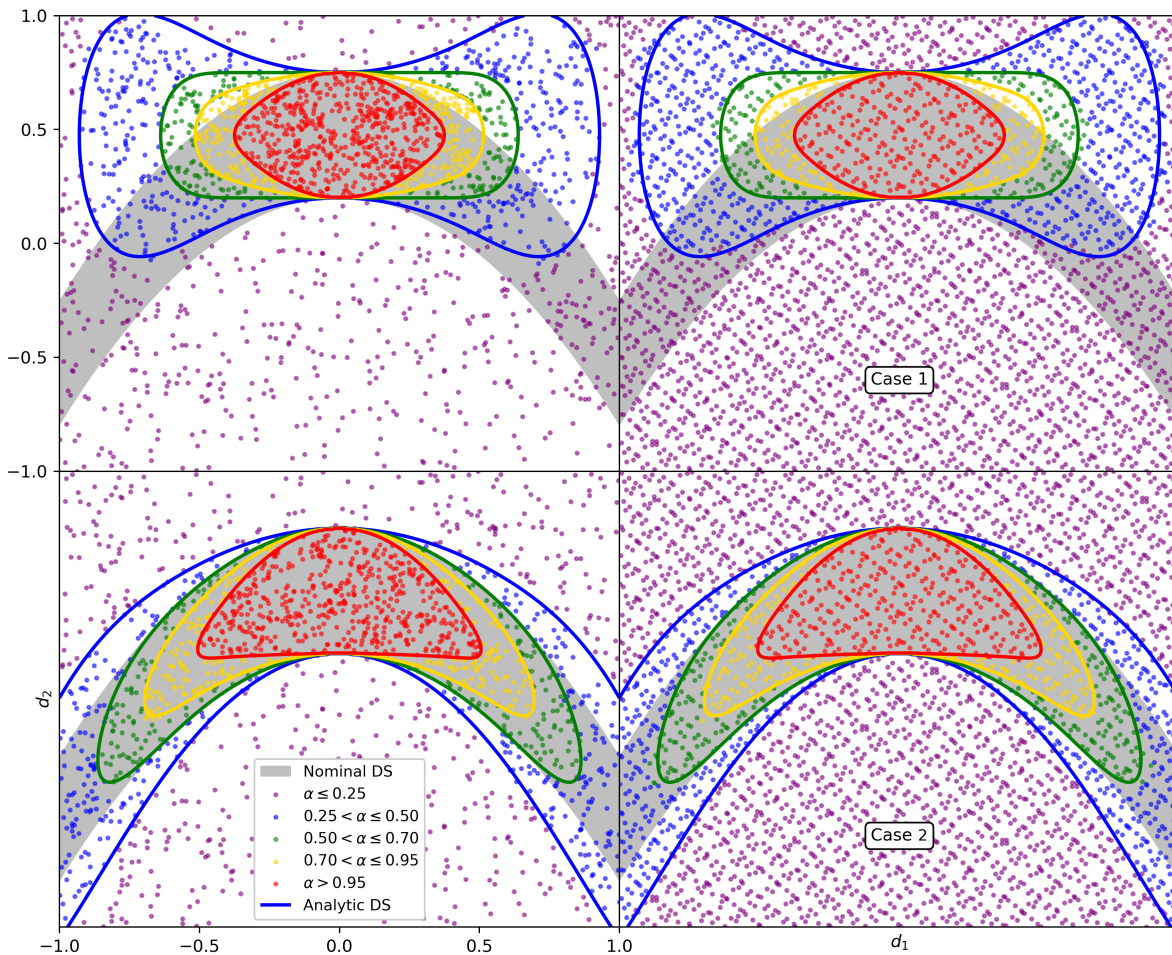
18

Figure 2: Comparison of sampling-based techniques to determine the probabilistic design space in Equation 10. The top and bottom plots are for the uncertainty scenarios $\theta \sim \mathcal{N}(0, 1)$ and $\theta \sim \mathcal{N}(1, \sqrt{0.3})$, respectively. The left plots show the results of nested sampling using Algorithm 2. The right plots show the results of the standard Monte Carlo in Algorithm 1 using quasi-random Sobol sampling. The nominal design space (Equation 9) is represented with the gray-shaded area. Iso-reliability contours for the actual probabilistic design space at different reliability values α (Equation 10) are shown in solid lines with different colors.
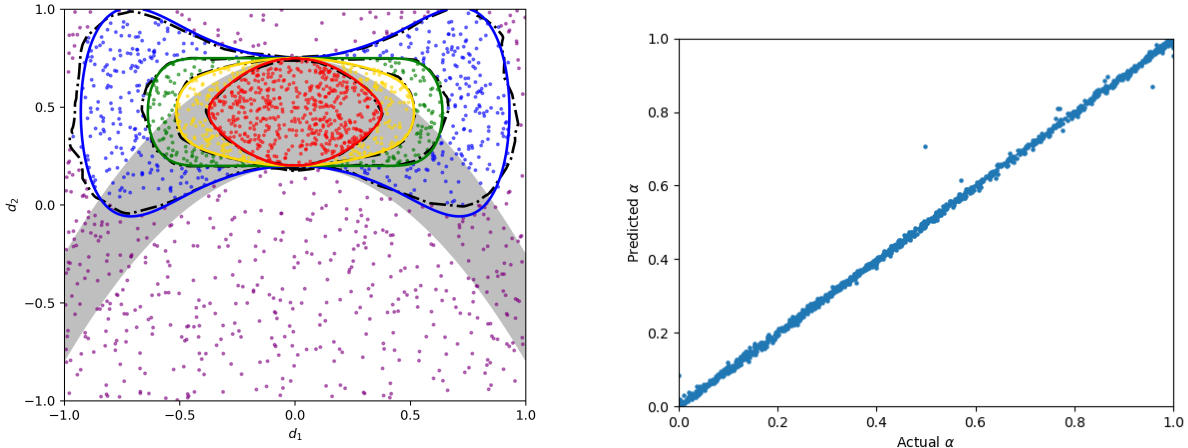
Figure 3: Exploitation of the nested sampling results using an MLP to fit the full feasibility probability map in the uncertainty scenario $\theta \sim \mathcal{N}(0, 1)$. The left plot compares the contours of the fitted MLP (dashed lines) with the actual DS (solid color lines) at different reliability values α. The gray-shaded area depicts the nominal design space (Equation 9). The right plot shows the parity plot of the trained MLP.

few samples being misclassified for the same reason as earlier with nested sampling since the same approach is used to estimate the feasibility probability. For a quantitative comparison between the two algorithms, notice that the total number of model evaluations is identical in both cases (322,000 and 369,200 model evaluations under the uncertainty scenarios $\theta \sim \mathcal{N}(0, 1)$ and $\theta \sim \mathcal{N}(1, \sqrt{0.3})$, respectively) and no acceleration strategy is applied for nested sampling. A breakdown of the samples generated with Algorithms 1 and 2 within different feasibility probability ranges is furthermore reported in Table 2. Recall that unsuccessful proposals in the nested sampling approach are currently discarded, which is the reason why the total number of samples is lower with this approach compared to Monte Carlo sampling. Regardless, nested sampling enables a much denser sampling of the targeted reliability region, as controlled by the number of live points and the gradual sampling towards higher reliability regions. It would take three to four times more samples—and hence model evaluations—for the standard Monte Carlo approach to achieve a similar concentration of points within the desired reliability region of 0.95 in this case study.

The left plot in Figure 3 compares several iso-contours of the exact feasibility probability

Table 2: Comparison between the samples generated via standard Monte Carlo with Sobol sampling (Algorithm 1) and nested sampling (Algorithm 2) within different reliability ranges.

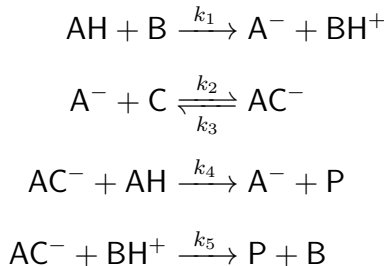| Reliability value	Samples drawn by standard Monte Carlo (Algorithm 1)	Samples drawn by nested sampling (Algorithm 2)
Uncertainty scenario: $\theta \sim \mathcal{N}(0,1)$		
$0.95 \leq \alpha$	247~~267~~	500
$0.70 \leq \alpha < 0.95$	139~~156~~	231~~281~~
$0.50 \leq \alpha < 0.70$	172~~212~~	181~~172~~
$0.25 \leq \alpha < 0.50$	666~~654~~	418~~388~~
$\alpha < 0.25$	2,025~~1,960~~	532~~524~~
Total	3,249	1,862~~1,865~~
Uncertainty scenario: $\theta \sim \mathcal{N}(1,\sqrt{0.3})$		
$0.95 \leq \alpha$	331~~354~~	500~~509~~
$0.70 \leq \alpha < 0.95$	194~~225~~	210~~259~~
$0.50 \leq \alpha < 0.70$	212~~256~~	200~~171~~
$0.25 \leq \alpha < 0.50$	490~~345~~	321~~204~~
$\alpha < 0.25$	2,022~~2,069~~	572~~597~~
Total	3,249	1,803~~1,740~~

map (Equation 9) with those of an MLP trained on the nested sampling results under the uncertainty scenario $\theta \sim \mathcal{N}(0,1)$. A good agreement is obtained in the upper reliability region where the concentration of samples is high, while the MLP predictions somewhat deviate in the lower reliability region where the sample points are scarcer—compare Table 2. The MLP under consideration here has four hidden layers, with 16, 32, 32 and 16 artificial neurons, respectively. Its training was conducted in the Python package Keras[41] that is implemented on top of the package TensorFlow[42], using mini-batch gradient descent. The resulting parity plot is shown on the right plot in Figure 3. The regression quality is generally good but for a few of the points that present a rather large deviation form the training set. Part of this mismatch is attributed to the fact that the data points in the training set itself are noisy due to the uncertainty discretization. Moreover, the actual feasibility probability map presents discontinuities at the points $\boldsymbol{d} = (0, 0.20)^\intercal$ and $\boldsymbol{d} = (0, 0.75)^\intercal$ so the MLP has difficulties capturing the very stiff probability variations near such points.

Industrial Case Studies

We demonstrate the applicability of the proposed nested sampling approach on two case studies of industrial relevance. The first one considers a Michael addition reaction in a continuous reactor and enables a comparison with recent work on flexibility-based algorithms[21] for probabilistic DS characterization. The second one investigates a biphasic Suzuki coupling reaction performed in fully batch mode[10] and gives the opportunity to determine a four-dimensional DS with a more complex mechanistic process model.

Michael Addition Reaction

This case study considers the following Michael addition reaction in a continuous stirred-tank reactor (CSTR):[21]

$$\mathsf{AH} + \mathsf{B} \xrightarrow{k_1} \mathsf{A}^- + \mathsf{BH}^+$$

$$\mathsf{A}^- + \mathsf{C} \underset{k_3}{\overset{k_2}{\rightleftharpoons}} \mathsf{AC}^-$$

$$\mathsf{AC}^- + \mathsf{AH} \xrightarrow{k_4} \mathsf{A}^- + \mathsf{P}$$

$$\mathsf{AC}^- + \mathsf{BH}^+ \xrightarrow{k_5} \mathsf{P} + \mathsf{B}$$

where AH is the Michael donor; C is the Michael acceptor; B is a base; BH^+, A^-, and AC^- are reaction intermediates; and P is the product. A complete statement of the steady-state process model can be found in Appendix B of the Supplementary Information. All of the kinetic constants are considered uncertain in this model, following a multivariate normal distribution $(k_1, \ldots, k_5) \sim \mathcal{N}(\mu_{\boldsymbol{k}}, \Sigma_{\boldsymbol{k}})$.

The knowledge space is defined in terms of two process parameters: (i) the molar ratio between the concentration of AH and B in the feed, $R_{\mathsf{AH}|\mathsf{B}} \in [10, 30]$; and (ii) the residence time in the CSTR, $\tau \in [400, 1400]$ (min). The feasible operating region is furthermore limited by two CQA constraints: (i) conversion of feed C greater than 90%; and (ii) residual

Table 3: Computational statistics for the Michael addition reaction using nested sampling (Algorithm 2) with different model uncertainty representations.

Uncertainty scenarios,	Nested sampling with acceleration		Nested sampling without acceleration	
N_θ	# model eval.	CPU sec.[1]	# model eval.	CPU sec.[1]
100	226,884	114	236,000	128
1,000	2,287,174	1,070	2,400,000	1,207

[1] CPU times obtained on a single core of AMD Ryzen 5 2600X processor.

concentration of AC^- smaller than $2\,\mathrm{mmol\,L^{-1}}$.

Like in the illustrative example above, we apply nested sampling (Algorithm 2) with a reliability target of $\alpha^* = 0.95$, now considering $N_\mathrm{L} = 1,000$ live points to enable a high density of points. The resulting sample set ($\mathcal{DS}$) is then regressed using an MLP comprising 4 hidden layers (with 16, 32, 32 and 16 neurons in each layer) to approximate the full feasibility probability map. The two plots in Figure 4 are for $N_\theta = 100$ and $1,000$ uncertainty samples. The corresponding computational statistics are reported in Table 3, both without and with the acceleration strategy. The effect of increasing the number of uncertainty scenarios by 10 fold is rather small on Figure 4, while the increase in CPU time is naturally close to a factor a 10 due to the corresponding increase in the number of function evaluations. Though differences can be noted between both fitted MLPs, with the MLP on the right plot ($N_\theta = 1,000$) predicting smoother iso-reliability contours due to a lower noise in the estimated feasibility probabilities compared to the MLP on the left plot ($N_\theta = 100$). It is also worth noting that the effect of the acceleration—about 5% reduction in the number of function evaluation and 10% reduction in CPU time—is rather modest. This is attributed to a relatively low fraction of rejected proposals due to the shape of the DS.

Finally, we compare the nested sampling results to the results obtained by Laky et al.[21] using an optimization approach based on flexibility analysis. It is worth reiterating that the algorithms therefrom determine or approximate the set-membership counterpart $\widehat{\mathcal{D}}_\alpha$ (Equation 6) to the probabilistic design space $\mathcal{D}_\alpha$, which happens to be quite conservative.[21] These
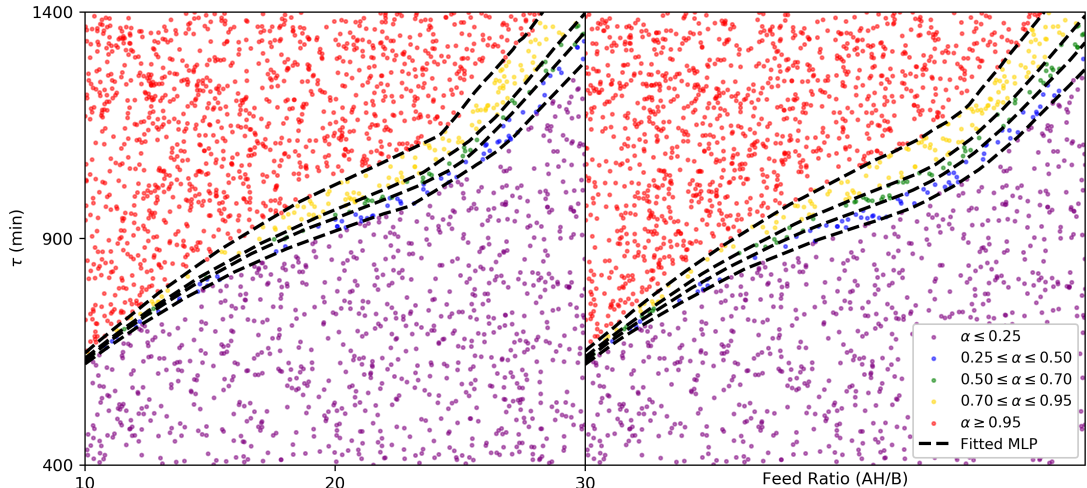
Figure 4: Comparison of probabilistic DS for the Michael addition reaction computed using nested sampling (Algorithm 2) with $N_\theta = 100$ (left) and $1,000$ (right) uncertainty scenarios. Samples from the probabilistic design space belonging to different reliability ranges are shown in different colors. Iso-reliability contours from the fitted MLP are shown in black dashed lines.

optimization-based methods are consistently faster than nested sampling (by a factor of 10 or more) when solving the optimization problems using a (local) gradient-based algorithm. But nested sampling is nevertheless competitive against these methods when a global optimization solver is used.

Suzuki Coupling Reaction

This case study investigates the Suzuki coupling reaction between a boronic ester (SM1) and an organohalide (SM2) to produce a desired pharmaceutical intermediate (P1) and a dimeric impurity (Imp1) related to SM1.[10] The reaction is biphasic and conducted in batch mode. The gaseous phase consists of an inert gas with traces of O_2; the liquid phase consists of a mixture of water and tetrahydrofuran (THF) as solvent for 17 chemical species that participate in 12 reactions—3 of which are reversible and 1 is considered instantaneous. A mechanistic, kinetic-based model is available describing the changes in composition of the liquid and gas phases during the batch—see Appendix C of the Supplementary Information

for a complete statement. The temperature-dependent reaction rates are modeled with Arrhenius equations and kinetic parameters that were verified experimentally.[10] The pre-exponential factors of all 14 kinetically-limited reactions are considered to be uncertain here, following normal distributions with standard-deviations equal to 15% of the nominal values.

The reactor has a large number of design and operation parameters so we consider a reduced DS characterization problem in terms of four parameters only: (i) the batch duration, $\tau \in [75, 300]$ (min); (ii) the equivalent of catalyst, $R_{\mathsf{Pd|SM2}} \in [0.001, 0.003]$ (mol mol^{-1}); (iii) the reactor temperature, $T \in [22, 64]$ (°C); and (iv) the molar fraction of O_2 in reactor's head space $y_{O_2} \in [10, 250]$ (ppm). Moreover, two CQAs limit the feasible operating region at the end of the batch: (i) maximum amount of unreacted $\mathsf{SM2}$ of 0.001 mol mol^{-1} for the reaction to be considered complete; and (ii) maximum level of impurity $\mathsf{Imp1}$ below 0.0015 mol mol^{-1} for the batch product to be downstream processable.

In order to determine the probabilistic DS we apply nested sampling (Algorithm 2) with a reliability target of $\alpha^* = 0.85$ ~~$\alpha^* = 0.95$~~. Because the DS now comprises four process parameters ~~we increase~~ the numbers of live points needs increasing to ensure high density of points, so we use ~~to~~ $N_{\mathrm{L}} = 5,000$ and $10,000$ next. We furthermore consider two uncertainty descriptions with $N_\theta = 200$ and $1,000$ samples, respectively. For the DS representation we use trellis charts where the ranges of oxygen concentration and temperature are split into four-by-four intervals—indicated by grey bars on the two outer axes; then each subplot is a projection of the points that belong to the particular intervals of oxygen fraction in head space y_{O_2} and temperature T on the plane defined by the batch duration τ and catalyst equivalent $R_{\mathsf{Pd|SM2}}$—the two inner axes. For instance, the chart on Figure 5 is a representation of the probabilistic DS computed with $10,000$ live points and $1,000$ ~~$N_\theta = 200$~~ uncertainty scenarios. Notice in particular the effect of the process parameters on the chart, whereby the DS expands upon increasing temperature at constant oxygen level and upon increasing oxygen fraction in head space at constant temperature. Computational statistics for the probabilistic DS are shown in Table 4. ~~using nested sampling with acceleration—~~ Overall, it takes over 4

Table 4: Computational statistics for the Suzuki coupling reaction using nested sampling (Algorithm 2) with different model uncertainty representations and live points.

Live Points,	Uncertainty scenarios,	Nested sampling with acceleration		Nested sampling without acceleration	
N_L	N_θ	# model eval.	CPU min.[1]	# model eval.	CPU min.[1]
5,000	200	1,412,390	525	1,451,200	574
10,000	1,000	14,132,975	6,464	14,544,000	6,741
5,000	1	–	–	6,392	3.5
10,000	1	–	–	12,680	6.8

[1] CPU times obtained on a single core of AMD Ryzen 5 2600X processor.

days~~over one day of CPU time~~ to determine the probabilistic DS with 10,000 live points and 1,000 uncertainty scenarios. A 12-fold decrease in the CPU time (to about 9 hours) is observed upon reducing to 5,000 live points and 200 uncertainty scenarios. The number of function evaluations is only reduced by about 3% in applying the acceleration strategy since the fraction of rejected proposals is again low. The corresponding reduction in CPU time is between 4–10%. A vectorized implementation of Algorithm 2 that would compute simulation ensembles or replacement proposals in parallel, could readily reduce the overall time needed and will be the focus of future work.

For comparison, we also compute the nominal DS (Equation 3) by applying nested sampling (Algorithm 2), in this case with a single uncertainty scenario ($N_\theta = 1$) corresponding to the nominal kinetic parameters and setting the reliability target to $\alpha^* = 1$. The chart in Figure 6 is a representation of this nominal DS with 10,000 live points. The results~~reported computational times~~ in Table 4 for 5,000 and 10,000 live points show a dramatic reduction in CPU time~~is dramatically reduced~~ to just a few minutes, which is clearly attributed to the fact that a single simulation is needed to assess whether or not a point belongs to the nominal DS instead of a large ensemble of simulations. However, the results in Figures 5 and 6 differ significantly as certain points within the nominal DS have a feasibility probability as low as 10–20%—an extremely poor reliability level in practice. Despite their popularity among practitioners approaches based on nominal parameter values or mean responses fail
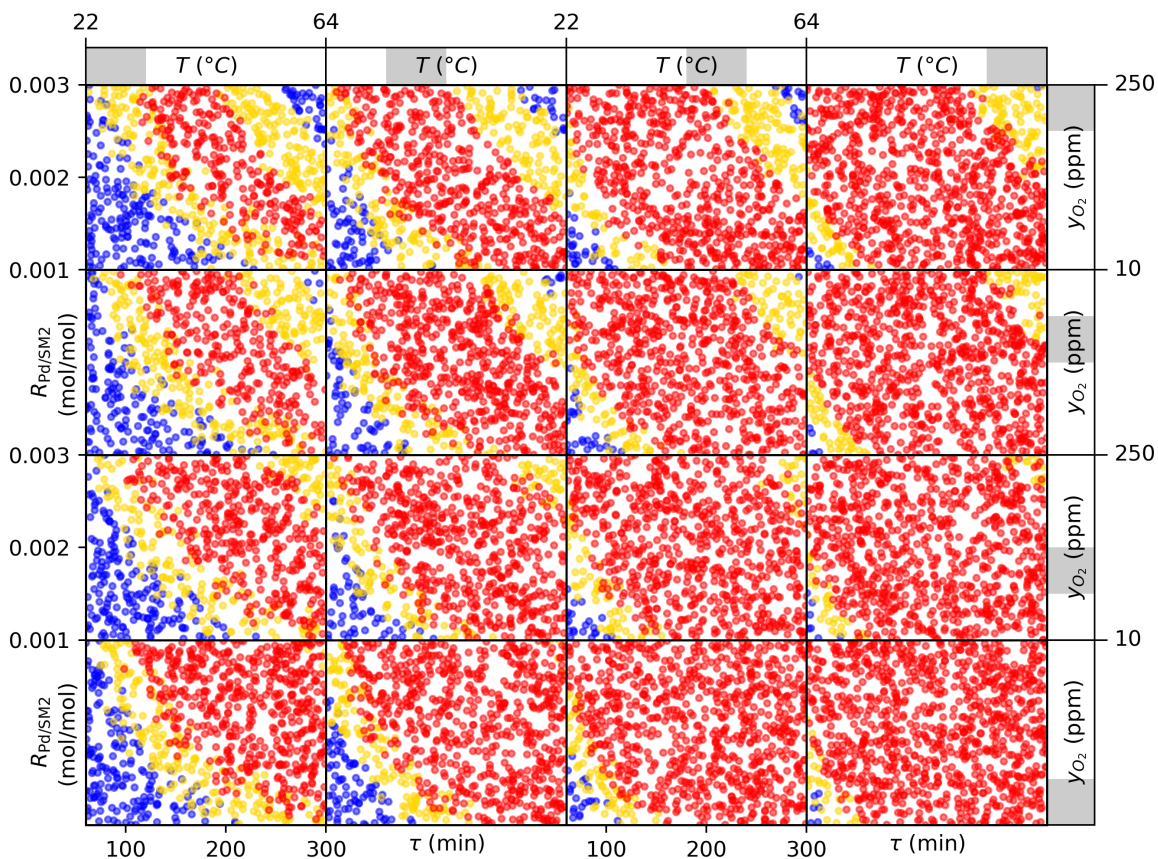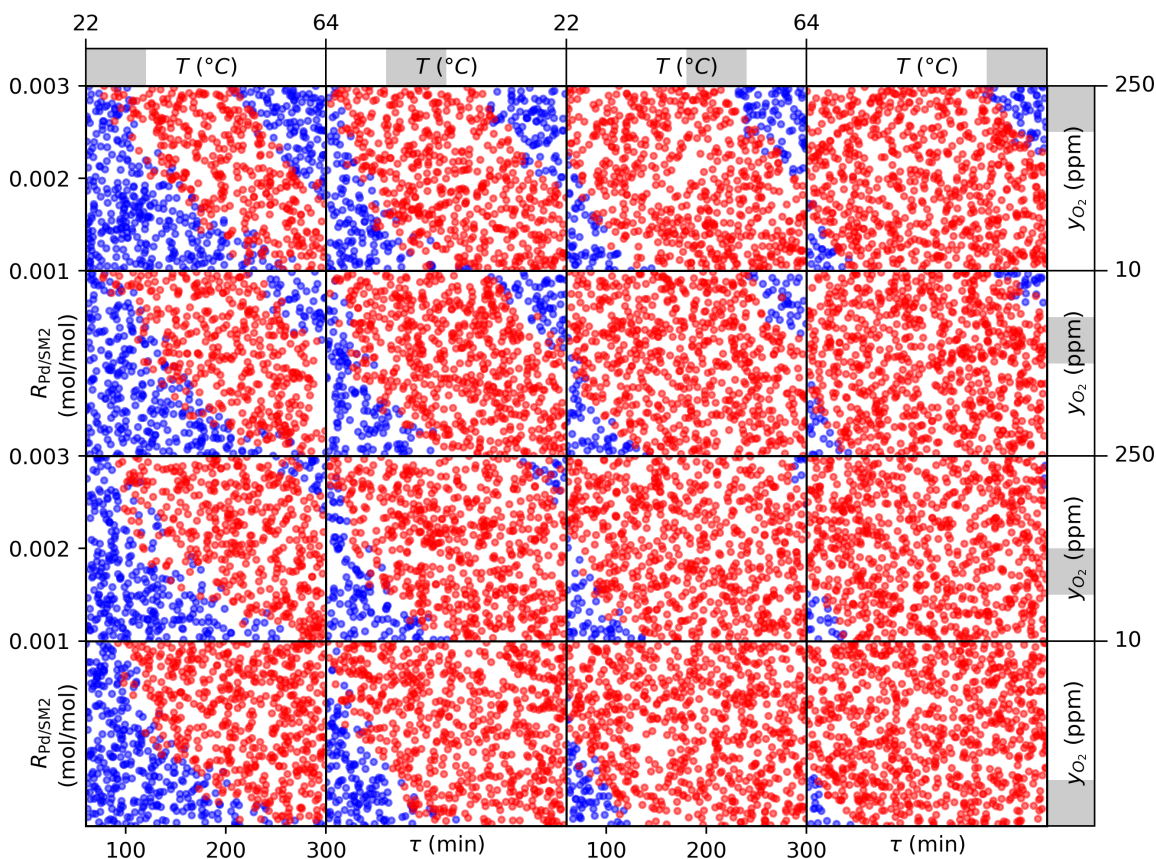
Figure 5: Probabilistic DS for the Suzuki coupling reaction computed using nested sampling (Algorithm 2) with $N_L = 10,000$ live points and $N_\theta = 1,000$ $N_\theta = 200$ uncertainty scenarios. The outer axes on the trellis chart correspond to the oxygen fraction in head space (y_{O_2}) and reactor temperature (T); the inner axes on each subplot, to the batch duration (τ) and catalyst equivalent ($R_{Pd|SM2}$). Samples from the probabilistic design space belonging to different reliability ranges are shown in different colors (red: $\alpha \geq 0.85$; yellow: $0.05 \leq \alpha < 0.85$; blue: $\alpha < 0.05$).

Figure 6: Nominal DS for the Suzuki coupling reaction computed using nested sampling (Algorithm 2) with $N_L = 10,000$ live points. The outer axes on the trellis chart correspond to the oxygen fraction in head space (y_{O_2}) and reactor temperature (T); the inner axes on each subplot, to the batch duration (τ) and catalyst equivalent ($R_{Pd|SM2}$). Samples within the nominal design space are depicted in red and those outside are in blue.

to quantify reliability and risk and, therefore, their results can be misleading. [10,22]

Concluding Remarks

Design space is a key concept in pharmaceutical QbD, helping practitioners develop a better understanding of their manufacturing processes and enhancing regulatory flexibility. In this context, it is of paramount importance to develop computational methods and tools that can provide a clear quantitative representation of the DS—in agreement with the ICH Q8

guideline. Our main focus throughout the paper has been on Bayesian approaches to DS characterization, which determine a feasibility probability that can be used as a measure of reliability and risk.

We have contributed a nested sampling algorithm tailored to the characterization of a probabilistic DS. A hallmark feature of nested sampling is its ability to maintain a given set of live points through regions with increasing probability feasibility until reaching a desired reliability level (or showing it cannot be reached). The algorithm furthermore leverages efficient strategies from Bayesian parameter estimation for generating replacement proposals during the search. Through a simple illustrative example and the case study of a Michael addition reaction, we have established that nested sampling can outperform standard Monte Carlo sampling and be competitive with optimization methods relying on process flexibility concepts, even in low-dimensional DS characterization problems. We have also showcased the use of machine learning techniques to reconstruct a feasibility probability map based on the sampled design space, which can be more easily exploited by the practitioner. In the second case study of a Suzuki coupling reaction we have shown that nested sampling is effective for larger DS characterization with a handful of process parameters, in the presence of a complex dynamic model and realistic model uncertainty—a class of problems currently out of reach for flexibility-based optimization techniques.

A major impediment facing nested sampling—and other sampling-based techniques—for probabilistic DS characterization in higher-dimensional problems, for instance in multi-unit integrated plants, is the very large number of process simulations required. This stems from the large ensemble of process simulations needed to capture the effect of model uncertainty in conjunction with maintaining a large number of live points. We have implemented an acceleration strategy as part of nested sampling, which reduces the overall number of process model runs without impairing a design space's accuracy. A recommended follow-up to this work would entail a vectorized implementation of the nested sampling algorithm in order to further reduce the time needed to characterize a complex DS, for instance by computing

29

simulation ensembles or replacement proposals in parallel.

Lastly, the characterization of the probabilistic DS in our work did not include adjustable control actions, although these may increase the size of the probabilistic design space. This practice is acceptable insofar as many pharmaceutical processes still use minimal online measurement and control of the CQAs—the process is instead carried to completion, followed by testing of the final product. As part of future work, it would be interesting to extend the nested sampling approach in order to encompass such recourse actions.

Acknowledgement

Data Statement: No new data was collected in the course of this research.

Supporting Information Available

The following supplementary information is available free of charge via the Internet at `http://pubs.acs.org/`.

- Appendix A: Nested Sampling for Bayesian Evidence

- Appendix B: Michael Addition Model

- Appendix C: Suzuki Coupling Model

- Appendix D: DEUS Python Package

References

(1) Troup, G. M.; Georgakis, C. Process systems engineering tools in the pharmaceutical industry. *Computers & Chemical Engineering* **2013**, *51*, 157–171.

(2) Reklaitis, G. V., Seymour, C., García-Muñoz, S., Eds. *Comprehensive Quality by Design for Pharmaceutical Product Development and Manufacture*; John Wiley & Sons, 2017.

(3) DiMasi, J. A.; Grabowski, H. G.; Hansen, R. W. Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics* **2016**, *47*, 20 – 33.

(4) ICH, Quality Guidelines. `https://www.ich.org/products/guidelines/quality/article/quality-guidelines.html`, 2019; accessed 15 August 2019.

(5) Holm, P.; Allesø, M.; Bryder, M. C.; Holm, R. In *ICH Quality Guidelines*; Teasdale, A., Elder, D., Nims, R. W., Eds.; John Wiley & Sons, 2017; Chapter 20, pp 535–577.

(6) Boukouvala, F.; Muzzio, F. J.; Ierapetritou, M. G. Design space of pharmaceutical processes using data-driven-based methods. *Journal of Pharmaceutical Innovation* **2010**, *5*, 119–137.

(7) Chatterjee, S.; Moore, C. M. V.; Nasr, M. M. In *Comprehensive Quality by Design for Pharmaceutical Product Development and Manufacture*; Reklaitis, G. V., Seymour, C., García-Muñoz, S., Eds.; John Wiley & Sons, 2017; Chapter 2, pp 9–24.

(8) Facco, P.; Dal Pastro, F.; Meneghetti, N.; Bezzo, F.; Barolo, M. Bracketing the design space within the knowledge space in pharmaceutical product development. *Industrial & Engineering Chemistry Research* **2015**, *54*, 5128–5138.

(9) Bano, G.; Facco, P.; Bezzo, F.; Barolo, M. Probabilistic Design space determination in pharmaceutical product development: A Bayesian/latent variable approach. *AIChE Journal* **2018**, *64*, 2438–2449.

(10) García-Muñoz, S.; Luciani, C. V.; Vaidyaraman, S.; Seibert, K. D. Definition of design spaces using mechanistic models and geometric projections of probability maps. *Organic Process Research & Development* **2015**, *19*, 1012–1023.

(11) Biegler, L. T.; Grossmann, I. E.; Westerberg, A. W. *Systematic Methods of Chemical Process Design*; Prentice Hall, 1997.

(12) Grossmann, I. E.; Calfa, B. A.; Garcia-Herreros, Evolution of concepts and models for quantifying resiliency and flexibility of chemical processes. *Computers & Chemical Engineering* **2014**, *70*, 22–34.

(13) Haleman, K. R.; Grossmann, I. E. Optimal Process Design Under Uncertainty. *AIChE Journal* **1983**, *29*, 425–433.

(14) Swaney, R. E.; Grossmann, I. E. An index for operational flexibility in chemical process design. Part I: Formulation and theory. *AIChE Journal* **1985**, *31*, 621–630.

(15) Georgakis, C.; Uztürk, D.; Subramanian, S.; Vinson, D. R. On the operability of continuous processes. *Control Engineering Practice* **2003**, *11*, 859–869.

(16) Lima, F. V.; Jia, Z.; Ierapetritou, M.; Georgakis, C. Similarities and differences between the concepts of operability and flexibility: The steady-state case. *AIChE Journal* **2010**, *56*, 702–716.

(17) Pantelides, C. C.; Shah, N.; Adjiman, C. S. Design Space, Models and Model Uncertainty. AIChE Annual Meeting. Nashville, TN, 2009.

(18) Straub, D. A.; Grossmann, I. E. Integrated stochastic metric of flexibility for systems

with discrete state and continuous parameter uncertainties. *Computers & Chemical Engineering* **1990**, *14*, 967–985.

(19) Pistikopoulos, E. N.; Mazzuchi, T. A. A novel flexibility analysis approach for processes with stochastic parameters. *Computers & Chemical Engineering* **1990**, *14*, 991–1000.

(20) Peterson, J. J. A Bayesian approach to the ICH Q8 definition of design space. *Journal of Biopharmaceutical Statistics* **2008**, *18*, 959–975.

(21) Laky, D.; Xu, S.; Rodriguez, J. S.; Vaidyaraman, S.; García Muñoz, S.; Laird, C. An Optimization-Based Framework to Define the Probabilistic Design Space of Pharmaceutical Processes with Model Uncertainty. *Processes* **2019**, *7*, 96.

(22) Peterson, J. J.; Yahyah, M.; Lief, K.; Hodnett, N. In *Comprehensive Quality by Design for Pharmaceutical Product Development and Manufacture*; Reklaitis, G. V., Seymour, C., García-Muñoz, S., Eds.; John Wiley & Sons, 2017; Chapter 4, pp 55–70.

(23) Straub, D. A.; Grossmann, I. E. Design optimization of stochastic flexibility. *Computers & Chemical Engineering* **1993**, *17*, 339–354.

(24) Ochoa, M. P.; Luciani, C.; Stamatis, S. D.; García-Muñoz, S.; Grossmann, I. E. New Developments in Flexibility Analysis in the Framework of Design Space Definition. AIChE Annual Meeting. Pittsburgh, PA, 2018.

(25) Floudas, C. A.; Gümüş, Z. H. Global Optimization in Design under Uncertainty: Feasibility Test and Flexibility Index Problems. *Industrial & Engineering Chemistry Research* **2001**, *40*, 4267–4282.

(26) Harwood, S. M.; Barton, P. I. How to solve a design centering problem. *Mathematical Methods of Operations Research* **2017**, *86*, 215–254.

(27) Boukouvala, F.; Muzzio, F. J.; Ierapetritou, M. G. In *Comprehensive Quality by Design for Pharmaceutical Product Development and Manufacture*; Reklaitis, G. V., Seymour, C., García-Muñoz, S., Eds.; John Wiley & Sons, 2017; Chapter 6, pp 95–123.

(28) Wang, Z.; Ierapetritou, M. G. Global sensitivity, feasibility, and flexibility analysis of continuous pharmaceutical manufacturing processes. *Computer Aided Chemical Engineering* **2018**, *41*, 189–213.

(29) Boukouvala, F.; Ierapetritou, M. G. Feasibility analysis of black-box processes using an adaptive sampling Kriging-based method. *Computers & Chemical Engineering* **2012**, *36*, 358–368.

(30) Rogers, A.; Ierapetritou, M. Feasibility and flexibility analysis of black-box processes. Part 1: Surrogate-based feasibility analysis. *Chemical Engineering Science* **2015**, *137*, 986–1004.

(31) Rogers, A.; Ierapetritou, M. Feasibility and flexibility analysis of black-box processes. Part 2: Surrogate-based flexibility analysis. *Chemical Engineering Science* **2015**, *137*, 1005–1013.

(32) Skilling, J. Nested Sampling. *AIP Conference Proceedings* **2004**, *735*, 395–405.

(33) Feroz, F.; Hobson, M. P.; Bridges, M. MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Monthly Notices of the Royal Astronomical Society* **2009**, *398*, 1601–1614.

(34) Mukherjee, P.; Parkinson, D.; Liddle, A. R. A nested sampling algorithm for cosmological model selection. *The Astrophysical Journal Letters* **2006**, *638*, L51.

(35) Handley, W.; Hobson, M.; Lasenby, A. PolyChord: nested sampling for cosmology. *Monthly Notices of the Royal Astronomical Society: Letters* **2015**, *450*, L61–L65.

(36) Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: New York, NY, 2004.

(37) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, 2009.

(38) Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, 2016.

(39) Schweidtmann, A. M.; Mitsos, A. Deterministic Global Optimization with Artificial Neural Networks Embedded. *Journal of Optimization Theory & Applications* **2019**, *180*, 925–948.

(40) Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Networks* **1991**, *4*, 251–257.

(41) Chollet, F.; et al., Keras. 2015; Software available from `https://keras.io`.

(42) Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mané, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viégas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; Zheng, X. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015; `http://tensorflow.org/`, Software available from tensorflow.org.

Graphical TOC Entry
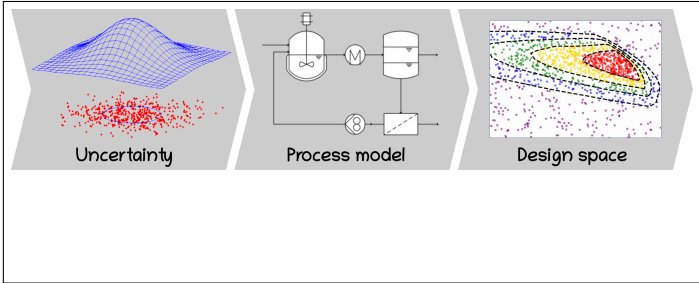
Nomenclature

Acronyms

CFD	Computational Fluid Dynamics
CPU	Central Processing Unit
CQA	Critical Quality Attribute
CSTR	Continuous Stirred-Tank Reactor
DS	Design Space
HPD	Highest Posterior Density
ICH	International Council for Harmonization
MCMC	Markov Chain Monte Carlo
MLP	Multi-Layer Perceptron
PSE	Process Systems Engineering
QbD	Quality by Design
QTPP	Quality Target Product Profile
DEUS	DEsign under Uncertainty using Sampling techniques

Main Symbols

α	reliability value
$\mathcal{D}$	design space
$\mathcal{K}$	knowledge space
$\mathcal{DS}$	sampled design space
$\mathcal{L}(\cdot)$	likelihood function
$\mathcal{N}(\mu, \sigma)$	Normal distribution with mean μ and standard deviation σ
$\mathcal{S}$	sampling set
$\mathbb{1}[\cdot]$	indicator function

$\mathbb{E}[\cdot]$	expected value
$\mathbb{P}[\cdot]$	probability
ω	weight
$\pi(\cdot)$	prior distribution
θ	model parameter
τ	residence time [min]
d	process parameter
G	critical quality attribute constraint
k	kinetic rate constant
N	number of samples
n	dimension
$p(\cdot)$	posterior distribution
R	molar ratio [$\mathrm{mol\,mol^{-1}}$]
s	critical quality attribute
T	temperature [°C]
y	gas molar fraction [ppm]
Z	evidence

Subscripts and Supercripts

$*$	target value
iter	iterations
L	live points
min	minimal value
nom	nominal
R	replacement proposals