

## Journal Pre-proof

Safe model-based design of experiments using Gaussian processes

P. Petsagkourakis, F. Galvanin

PII: S0098-1354(21)00117-4  
DOI: <https://doi.org/10.1016/j.compchemeng.2021.107339>  
Reference: CACE 107339

To appear in: *Computers and Chemical Engineering*

Received date: 27 November 2020  
Revised date: 5 March 2021  
Accepted date: 13 April 2021

Please cite this article as: P. Petsagkourakis, F. Galvanin, Safe model-based design of experiments using Gaussian processes, *Computers and Chemical Engineering* (2021), doi: <https://doi.org/10.1016/j.compchemeng.2021.107339>



This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier Ltd.

**Highlights:**

- Safe application of experimental design using Gaussian processes.
- Combination of trust-region ideas to balance exploration vs exploitation.
- Surrogate based approximation for Fisher information matrix.
- Construction of hybrid model for constraint satisfaction.

Journal Pre-proof

## Safe model-based design of experiments using Gaussian processes

P. Petsagkourakis<sup>a</sup>, F. Galvanin<sup>a,\*</sup><sup>a</sup>Centre for Process Systems Engineering (CPSE), Department of Chemical Engineering, University College London, Torrington Place, London WC1E 7JE, United Kingdom

---

**Abstract**

The construction of kinetic models has become an indispensable step in developing and scale-up of processes in the industry. Model-based design of experiments (MBDoe) has been widely used to improve parameter precision in nonlinear dynamic systems. Such a framework needs to account for both parametric and structural uncertainty, as the physical or safety constraints imposed on the system may well turn out to be violated, leading to unsafe experimental conditions when an optimally designed experiment is performed. In this work, Gaussian processes are utilized in a two-fold manner: 1) to quantify the uncertainty realization of the physical system and calculate the plant-model mismatch, 2) to compute the optimal experimental design while accounting for the parametric uncertainty. TheOur proposed method, Gaussian process-based MBDoe (GP-MBDoe), guarantees the probabilistic satisfaction of the constraints in the context of the model-based design of experiments. GP-MBDoe is assisted with the use of adaptive trust regions to facilitate a satisfactory local approximation. The proposed method can allow the design of optimal experiments starting from limited preliminary knowledge of the parameter set, leading to a safe exploration of the parameter space. This method's performance is demonstrated through illustrative case studies regarding the parameter identification of kinetic models in flow reactors.

---

**1. Introduction**

Biochemical, physicochemical or catalytic processes are often too complex to develop accurate physics-based only models, and plant-model mismatch is inevitable. It is common to develop approximate kinetic models under some assumptions regarding the system's mechanism to represent the physical system accurately. The availability of a trustworthy kinetic model could be used to predict the system's behaviour outside of the validated experimental conditions and then be employed for design, optimization and control in process systems engineering applications [6]. The model identification requires experimental data to evaluate the validity of a proposed model and estimate the model parameters to match the physical system's behaviour over the selected range of operating conditions. Nevertheless, the experimental procedure may become a costly, very time consuming and infeasible task when poor experimental designs are implemented. Identifying an appropriate approximated model is highly dependent on limitations due to the certain observability of the physical phenomena. Additionally, the respected data may be difficult and expensive to obtain because of physical accessibility and/or limitations on the experimental budget.

Model-based design of experiments (MBDoe) has been widely utilized to improve parameter precision in highly nonlinear dynamic systems. The main goal of MBDoe is to design control inputs and sampling schedules

---

\*Corresponding author

Email address: [f.galvanin@ucl.ac.uk](mailto:f.galvanin@ucl.ac.uk) (F. Galvanin)

[17] to achieve the most informative experiment possible by satisfying constraints on feasibility and safety of operation. The optimal MBDoE problem for improving parameter estimation involves the maximization of a measure of the expected information (or the minimization of a measure of the expected variance of model parameters) by acting on experiment decision variables to ensure the feasibility of the experimental trials. However, since the optimization problem is based on the available model, both plant-model mismatch and parametric uncertainty may significantly affect the result [43, 37].

The literature consists of various approaches to ensure optimality and feasibility, particularly under the presence of parametric uncertainty [45]. Two distinctly different scenarios are considered regarding the information and the constraint satisfaction, robust and stochastic approaches. A robust approach for the information content was proposed [4], where the predicted information content of experiments needs to be optimal for the whole space of the uncertain parameters. The robust optimization problem is usually solved in terms of min-max optimization, in which uncertainties are typically assumed to be deterministic and bounded [45, 30, 66, 53]. This problem is often numerically intractable when candidate models are highly nonlinear. Various methods have been proposed to accommodate this limitation, including linearization of the inner optimization [30] and a sensitivity-based approximate robust approach assuming ellipsoidal joint confidence region model parameters [35]. The robust approaches are typically conservative as they require the optimality and satisfaction of constraints to hold for all the possible values within the predefined uncertainty region. On the contrary, a stochastic approach has been popularized, where the information content is maximized in expectation, and the constraints are satisfied with a given probability [4]. The stochastic (also called probabilistic) experimental design framework avoids the typical strong conservatism of the worst-case MBDoE techniques, as the probability of occurrence of different realizations is accounted for differently. Various approaches [61, 35, 19, 58, 33] have been proposed in this family of problems usually inspired by the techniques originated from optimal control problem.

The satisfaction of chance constraints has been widely studied in the field of optimal control, especially from the stochastic model predictive control community (SMPC). Comprehensive reviews regarding SMPC can be found in [32, 16, 21]. Additionally, both parametric and structural mismatches have been incorporated in [1, 25, 3]. Probabilistic (chance) constraints are incorporated into the MBDoE problem to seek a trade-off between a designed experiment's information content and allow for a user-defined level of risk during the experiment. Most approaches compute the expected information given the (previous) parameter identification procedure, where confidence regions are constructed using approximation techniques [40] around the point estimates for the parameters.

Galvanin et al. [19] directly utilized a Monte Carlo approach to compute the constraint tightening (backoff) given the approximated probability distribution of the parameters. Such a method is very accurate for estimating the constraint tightening for the user-defined constraint satisfaction but computationally costly. Telen et al. [61] proposed the use of the unscented transformation [64] to propagate the parametric uncertainty in the objective function and the constraints. The unscented transformation has widely been utilized in the optimal control [7] due to the low computational complexity. Nevertheless, unscented transformation approaches may fail in the presence of strong nonlinear relationships between the parameters and the states of the model. In this context, polynomial chaos expansion (PCE) has been used to compute the required moments of the objective and constraints in [33] and recently in [35]. Polynomial chaos expansions have been used in stochastic model predictive control for the purposes of uncertainty propagation through a nonlinear model [67]. Similarly,

Gaussian process (GP) [51, 65] has been proposed as an alternative to PCE, as they are non-parametric models. Gaussian processes have been employed due to the estimation of uncertainty of prediction around points caused by an insufficient amount of data-points. The GP usually assumes an independent and identically distributed (iid) Gaussian noise added in the observed measurements, and the outputs obtained at different inputs are jointly Gaussian distributed, where a kernel function defines the covariance. Gaussian process has not found a significant application in MBDoE, with exception in [36] in the context of model discrimination. In this case the GP is used to approximate the different models and marginalize the parameters of the models given the computed uncertainty. Nevertheless, GP has extensively been used in the area of optimal control, where the prior models and black-box techniques are combined to describe the system. GP may be coupled with approximate method for propagating the uncertainty as unscented Kalman Filter [29], exact moment matching [22, 14], linearization [22] and scenario-based approaches [63, 9]. Recently, GPs have been used in hybrid modeling rational, where they are coupled with the prior (nominal) model that is assumed for the system, and learning is conducted only for the GP [15, 24].

The methods of robust MBDoE, e.g. [4, 35] use the confidence regions computed in the parameter estimation step. Such an assumption may be sufficient to accommodate the uncertainty for the objective of the information content. However, constraint satisfaction should require different treatment. The satisfaction of the constraints is typically guaranteed given the uncertainty of the parameters and the required assumptions, including the availability of a large number of available prior experimental data points. In the typical scenario of paucity of data at the beginning and a presence of model mismatch, these assumptions do not hold. As the methodology is model-based, both model mismatch (i.e., a model structure inadequate to represent the physical systems) and parametric mismatch (i.e., incorrect values of the parameters) may affect the consistency of the whole design procedure [40, 44]. These assumptions have been avoided in some works using a disturbance based estimation [20] and the MBDoE in the context of guaranteed parameter estimation [52]. Additionally, Quaglio et al. [46] developed a method to find the region where the approximated model is valid and avoid the intrinsic issues with the model-mismatch in MBDoE methods [47].

In this work, we propose a novel MBDoE framework suitable for guaranteeing operation feasibility when a plant-model mismatch is strongly present, i.e. possible disturbances/hardware malfunctions and the initial limited data. The plant-model mismatch is a critical issue in optimal control, where stability [42, 41], offset-free tracking [39] and satisfaction of constraints has been extensively studied [24].

We treat the objective function of the MBDoE by utilizing a stochastic surrogate, i.e. GP. Additionally, we focus on the satisfaction of constraints that has not been addressed in the literature for the optimal experimental designs avoiding the use of covariance of the parameters that have been computed from the parameter estimation. Gaussian process is utilized to compute the model mismatch together with a trust region around operational point, that is updated iteratively. This is inspired by the derivative-free methods [13] that were recently applied in [15]. This way, restrictive assumptions of the nature of the ‘real’ model are avoided.

The structure of this paper is as follows. The problem definition is outlined in section 2, the details of the proposed method for safe, optimal experimental design is presented in section 3. A case study is presented in section 4, where the framework is applied to two separate case studies, and in the last section, conclusions are outlined.

## 2. Problem Definition

We assume that the process under consideration can be represented by a set of differential and algebraic equations (DAEs), where measured variables  $\mathbf{y}$  can only be measured in finite sampling times  $\{1, \dots, N\}$

$$\begin{aligned} \mathbf{f}(\dot{\mathbf{x}}, \mathbf{x}, \mathbf{u}, \mathbf{w}, \boldsymbol{\theta}^*) &= 0 \\ \mathbf{y} &= \mathbf{h}(\mathbf{x}, \mathbf{u}, \boldsymbol{\nu}), \end{aligned} \quad (1)$$

where  $\mathbf{x} \in \mathbb{R}^{n_x}$ ,  $\mathbf{u} \in \mathbb{R}^{n_u}$ ,  $\mathbf{y} \in \mathbb{R}^{n_y}$  and  $\boldsymbol{\theta}^* \in \mathbb{R}^{n_{\theta^*}}$  are the vector of state variables representing the true system, vector of control (manipulated) variables, vector of measured (output) variables and parameters of the process, correspondingly. Additionally, the physical system is assumed to be subject to additive normally distributed disturbance ( $\mathbf{w} \in \mathbb{R}^{n_w}$ ) and noise ( $\boldsymbol{\nu} \in \mathbb{R}^{n_\nu}$ ). The functions  $\mathbf{f} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_w} \rightarrow \mathbb{R}^{n_x}$  and  $\mathbf{h} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_w} \rightarrow \mathbb{R}^{n_y}$  are unknown functions that describe the observed process. For practical reasons this model will be referred as ‘true model’.

In practice, the mapping relating the process inputs and outputs of (1) is typically unknown, and only an approximate model is available, parametrized by a set of parameters  $\boldsymbol{\theta}$ . However, in many cases [43, 2] it is unlikely that a model’s predictions coincide with those of the real system due to the relevant assumption, like the possible plant-model mismatch and/or equipment malfunction (e.g. bias in the pump, wrongly calibrated HPLC). Let the approximated model derived from simplifying hypotheses and/or ignoring equipment’s malfunctions be:

$$\begin{aligned} \hat{\mathbf{f}}(\hat{\mathbf{x}}, \hat{\mathbf{x}}, \mathbf{u}, \boldsymbol{\theta}) &= 0 \\ \hat{\mathbf{y}} &= \hat{\mathbf{h}}(\hat{\mathbf{x}}, \mathbf{u}), \end{aligned} \quad (2)$$

where  $\hat{(\cdot)}$  represents the model predictions. The expected values of the parameters  $\boldsymbol{\theta}$  and their corresponding distribution are estimated given available measurements. Notice that the vector of state variables for the model ( $\hat{\mathbf{x}}$ ) is not necessarily the one from (1), as different models may have different states to describe the behaviour.

Optimal model-based design of experiments (MBD<sub>oE</sub>) for parameter estimation is employed to reduce the variance of the parameters  $\boldsymbol{\theta}$ , which is part of the model identification procedure. This technique is usually embodied in a sequence of three key activities: experimental design (MBD<sub>oE</sub>), experimental execution and parameter estimation. Such framework is particularly flexible, allowing for the definition of a set of active constraints on both state and design variables during the optimization. At the experimental design step, MBD<sub>oE</sub> is usually inferred as an optimization problem in which the optimal experimental conditions are determined (control variables  $\mathbf{u}$ , sampling times, duration of experiments) to minimize a measure of the predicted covariance matrix or maximize a measure of the expected Fisher information matrix (FIM). Most importantly, the designed experiments should take into account feasibility specification (e.g. safety constraints) during the experimental trials.

### 2.1. Maximum Likelihood Estimation & Confidence Regions

The uncertainty of the model parameter is computed using the maximum likelihood estimation results, and the respective confidence regions are used for the robust/stochastic optimally informative experiment or robust/chance constraint satisfaction. In this subsection, the most common methods to construct confidence regions are discussed, which are usually used for addressing the feasibility problem of chance constraints. The limitation of such approaches is also given.

Let  $n_b$  be the experiments with  $n_m$  measurements be available  $\mathcal{D} = \{\mathbf{y}_b(t_m), \mathbf{u}_b(t_m), \mathbf{x}_m(0), \forall b, m \in n_b, n_m\}$ . If the manipulated variables are assumed to be precisely known, then the joint probability of the prediction-observation mismatch in the available measurements for the parameter values  $\boldsymbol{\theta}$  is the following log-likelihood:

$$-2 \log(\mathcal{L}(\boldsymbol{\theta}|\mathcal{D})) = \log(2\pi^{n_b n_m}) + \sum_{b=1}^{n_b} \sum_{m=1}^{n_{meas}} \log(|\boldsymbol{\Sigma}_{m_i}^y|) + (\hat{\mathbf{y}}_b(t_m, \boldsymbol{\theta}) - \mathbf{y}_b(t_m))^T \boldsymbol{\Sigma}_{b_m}^y (\hat{\mathbf{y}}_b(t_m, \boldsymbol{\theta}) - \mathbf{y}_b(t_m)), \quad (3)$$

where  $\hat{\mathbf{y}}_b(t_m, \boldsymbol{\theta})$  corresponds to the solution of (2). The maximum-likelihood estimation computes  $\hat{\boldsymbol{\theta}}$  to maximize the likelihood function ( $\mathcal{L}$ ) or minimize (3) subject to (2). Confidence regions can be computed, under the assumption that model (2) is correct and the respected computed parameters parameters ( $\hat{\boldsymbol{\theta}}$ ) are close to the (unique) ‘true’ values of the model parameters ( $\boldsymbol{\theta}$ ). Then for a large number of available measurements, both the likelihood subset ratio statistic and the Wald subset statistic follows a chi-square distribution [56, 31]. These asymptotic confidence results can be used to obtain 100(1 -  $\alpha$ )% likelihood-based,  $\Theta_{\mathcal{L}}$ , (4) and  $\Theta_{\mathcal{W}}$  Wald (5) confidence regions:

$$\Theta_{\mathcal{L}} := -2 \log \left[ \frac{\mathcal{L}(\boldsymbol{\theta}|\mathcal{D})}{\mathcal{L}(\hat{\boldsymbol{\theta}}|\mathcal{D})} \right] \leq \chi_{n_{dof}}^2(1 - \alpha), \quad (4)$$

$$\Theta_{\mathcal{W}} := (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{V}_{\boldsymbol{\theta}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq \chi_{n_{dof}}^2(1 - \alpha), \quad (5)$$

with  $\mathbf{V}_{\boldsymbol{\theta}}$  being the covariance for the parameters calculated at  $\hat{\boldsymbol{\theta}}$  and  $\chi_{n_{dof}}^2(1 - \alpha)$  is the (1 -  $\alpha$ ) quantile of the chi-square distribution with  $n_{dof}$  degrees of freedom. The covariance  $\mathbf{V}_{\boldsymbol{\theta}}$  is usually approximated by a Taylor expansion of the likelihood. Particularly, (5) has extensively been utilized in the design of experiments [35, 4] as the closed-form expression is available compare to the likelihood-subset ratio (4), where there is no closed-form expression available [54]. Nevertheless, the Wald confidence regions depend on the model reparameterization because of the (approximate) covariance term  $\mathbf{V}_{\boldsymbol{\theta}}$ , which may significantly affect the resulted confidence regions [50]. Additionally, in the same rationale, a Laplace approximation can be used for the posterior distribution of the parameter estimates given the available data:

$$p(\boldsymbol{\theta}|\mathcal{D}) \sim \mathcal{N}(\hat{\boldsymbol{\theta}}, \mathbf{V}_{\boldsymbol{\theta}}), \quad (6)$$

where  $\mathbf{V}_{\boldsymbol{\theta}}$  is the covariance for the parameters at  $\hat{\boldsymbol{\theta}}$ , similar to (5). The parametric uncertainty estimation of (4-6) are the most common used strategies to approximate confidence regions or posterior distribution of the parameters. Two additional methods can be used that are more computationally demanding: Bayesian-based estimation [23] and set-membership estimation [40].

In the context of stochastic MBDoE, these strategies are used to exploit (4-6) to formulate a stochastic optimization problem subject to chance constraints. The parametric uncertainty estimation given by (4-6) is affected by several limitations:

- i) A large amount of data is required. However, this is not the case when we design experiments since we want to keep the number of data-points to the minimum.
- ii) The covariance matrix  $\mathbf{V}_{\boldsymbol{\theta}}$  is approximated by a Taylor expansion. This means that the highly nonlinear models will result in unreliable covariance functions.
- iii) The expression in (4) is more accurate but with disadvantages (see point (i)) and there is no closed-form expression. Hence, Monte Carlo simulations or other numerical solutions [54] are required.

### 2.2. General Problem Formulation

The general formulation of the problem is constructed here where 1) the constraints are satisfied for a true model, and 2) the new experiment is the most informative for a given approximated model. The following infinite-dimensional problem can describe the proposed experimental design,  $\mathcal{P}(\mathbf{u})$  with  $\mathbf{u}$  being the optimization variables:

$$\mathcal{P}(\mathbf{u}) := \left\{ \begin{array}{l} \max_{\mathbf{u}} \mathbb{E}_{\boldsymbol{\theta} \sim \Theta} \left( \psi \left( \frac{\partial \hat{\mathbf{x}}(t_N)}{\partial \boldsymbol{\theta}}^\top \Sigma_{exp} \frac{\partial \hat{\mathbf{x}}(t_N)}{\partial \boldsymbol{\theta}} + \mathbf{M}_0 \right) \right) \\ \text{s.t.} \\ \text{Approximated Model:} \\ \hat{\mathbf{x}}(t_0) = \mathbf{x}(0) \\ \hat{\mathbf{f}}(\dot{\hat{\mathbf{x}}}, \hat{\mathbf{x}}, \mathbf{u}, \boldsymbol{\theta}) = 0 \\ \hat{\mathbf{y}} = \hat{\mathbf{h}}(\hat{\mathbf{x}}, \mathbf{u}) \\ \text{Real System:} \\ \mathbf{x}(t_0) = \mathbf{x}(0) \\ \mathbf{f}(\dot{\mathbf{x}}, \mathbf{x}, \mathbf{u}, \boldsymbol{\nu}) = 0 \\ \mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{u}, \boldsymbol{\nu}) \\ \text{Constraints:} \\ \mathbf{u} \in \mathbb{U} \\ \mathbb{P}(g_j(\mathbf{x}, \mathbf{u}) \in \mathbb{X}) \geq 1 - \bar{\alpha}, \forall j \in \{1, \dots, n_g\}, \end{array} \right. \quad \begin{array}{l} (7) \\ \\ \\ (8) \\ \\ \\ (9) \\ \\ (10) \end{array}$$

where  $\psi(\cdot)$  is a scalar function of the Fisher information (it depends on the selected design criterion, e.g. A-optimal, D-optimal etc), that is maximized in expectation given  $\Theta$  and  $\mathbf{M}_0$  is the prior Fisher information matrix representing the prior knowledge on the parametric system as derived from preliminary experiments. Additionally, the equality path constraints (8) corresponds to the DAE that describes the approximated model parametrized by  $\boldsymbol{\theta}$  and characterized by states  $\hat{\mathbf{x}}$ , the chance constraints (10) are subject to the real behaviour of the physical system that in this case is represented as a stochastic dynamic system (9) with states  $\mathbf{x}$ . Additionally, the constraints are meant to be satisfied with probability higher than  $1 - \bar{\alpha} \in [0, 1]$ , a user-defined parameter representing the probability of constraint satisfaction with values close to 1.

### 2.3. Standard Framework

In this section the standard framework is described. First, each stage of the standard framework for MBDoE is briefly given.

A block diagram with the standard framework is given in Fig. 1a. The fundamental steps are briefly illustrated:

1. A preliminary set of data is imported in the system.
2. Parameter estimation is performed for the approximated model.
3. The estimated parameters are passed in (cyan) design block, where the new experimental conditions are computed. This step is the core step, where the uncertainty of the parameters together with their corresponding estimates are used to estimate the following experimental conditions.



4. The experiments are performed and the data-set is updated

This closed-loop is terminated when any of the following holds:

- Statistics are satisfied.
- The optimal conditions do not change in between iterations.

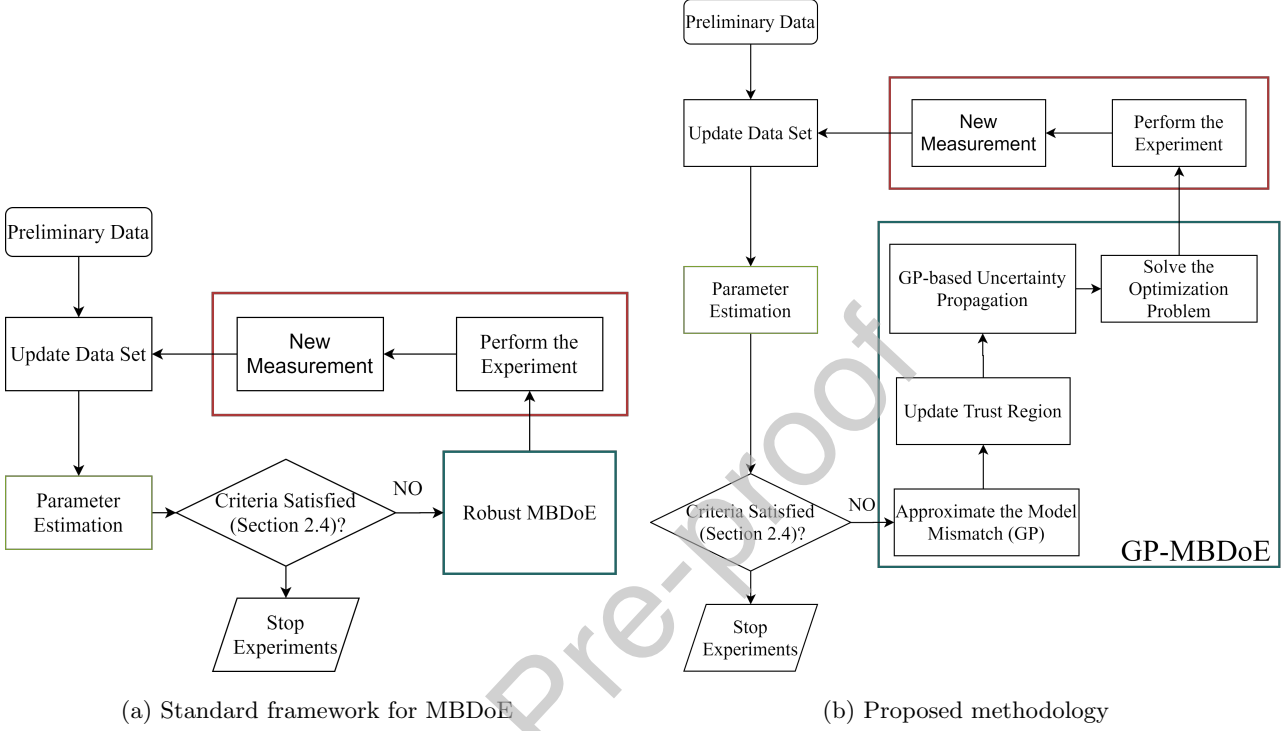


Figure 1: Standard (a) and proposed (b) frameworks are presented here, where the standard methodology performs parameter estimation and then with the update kinetic parameters designs the next experiment. On the other hand, the proposed methodology aims to construct and update trust regions and GP to ensure the chance constraints satisfaction.

#### 2.4. Termination Criteria.

The procedures described in Fig. 1 may terminate when either of the following two conditions are met: i) The design variables do not change within two cycles or ii) the statistics are satisfied. The parameters  $\hat{\theta}$  and the model predictions that are obtained from the parameter estimation step are associated with the  $t$ -values and the sum of squared residuals,  $\chi_{sample}^2$ :

$$\chi_{sample}^2 = \frac{(y_{ij} - \hat{y}_{ij})^2}{\sigma_{y_i}^2} \quad (11)$$

$$t_j = \frac{\hat{\theta}_j}{\sqrt{\mathbf{V}_{\theta_{jj}} t_{ref}(1 - \alpha/2, N_y - N_\theta)}}$$

where  $t_{ref}$  is a reference t-value given by a Student t-distribution with  $N_y - N_\theta$  degrees of freedom. The model provides an accurate representation of experimental data if  $\chi_{sample}^2 < \chi^2(1 - \alpha, N_y - N_\theta)$ , with  $\chi^2(1 - \alpha, N_y - N_\theta)$  a reference value from a  $\chi^2$  distribution with  $N_y - N_\theta$  degrees of freedom and confidence level  $1 - \alpha$ . The model parameter  $j$  requires  $t_j > t_{ref}(1 - \alpha, n_y - n_\theta)$  to be statistically significant.

### 3. Methodology

This section presents the proposed methodology. Gaussian process is used in a two-fold manner: 1) to represent the model mismatch between the available model and the collected data is approximated using GP. To ensure the reliability of the predictions of the GP, trust regions will be applied to constrain the region of experimental conditions where predictions are reliable. 2) An additional GP is then employed to compute the expected value and variance of the objective (i.e. metric of expected information/variance?) with respect to the parametric uncertainty.

The intuition behind the two-fold use of GPs is the non-parametric approximation for the physical system to enforce the constraint satisfaction and the same time, maximize the information of the approximated parametric model in expectation. Notice that a global GP would not be sufficient since it suffers from the same issues that the approximated confidence regions have; it requires a sufficient number of data-points. If there are enough data-points, the GP can provide a prediction for both the mean and the variance for the approximation. In the presence of limited data points, the GP may result in unreliable predictions unless assumptions are made regarding the function's complexity (see section 3).

#### 3.1. Proposed framework

An 'ideal' methodology would solve the optimization problem described in (7-10). However such an approach is intractable for the following reasons: 1) The real system is not known and 2) the expectation and probabilities are in general intractable integrals. To accommodate these issues, we propose a novel methodology. A block diagram with the proposed framework is given in Fig. 1b. The robust MBD<sub>oE</sub> block in the standard framework (1a) is replaced by theour proposed method, termed Gaussian process-based MBD<sub>oE</sub> (GP-MBD<sub>oE</sub>)

The proposed algorithm is based on GP to approximate the model mismatch between the process and the model, additional trust regions are applied to restrict the feasible design space per iteration. Trust regions are additive constraints for the design variables to limit the distance of the next design from the previous design. After that an additional GP is employed in Fig. 1b block 'GP-based Uncertainty Propagation' to approximate the expectation of the metric of the Fisher information and propagate the posterior distribution of the parameters.

The GP-MBD<sub>oE</sub> performs 4 additional steps before the computation of the new design.

- With the available data, a GP is computed for the plant-model mismatch (details given in Section 3.3).
- A trust region (TR) is updated-computed to restrict the new designs and allows the use of GP locally (details given in Section 3.4).
- The parametric uncertainty (e.g. confidence regions) can be computed via the methods described in section 2.1 in the parameter estimation stage. The GP at this stage aims to approximate the input-output behaviour of the parametric model and not the real system; hence in-silico data points are only employed. The parametric uncertainty is then propagated to compute the objective function of the design of experiments (more details are given in Section 3.5 and Appendix A and in [36]).
- The above are then used to perform the optimization step and find the new optimal design with the safety considerations included (details given in Section 3.6).

### 3.2. Gaussian Process Fundamentals

In this section GP is introduced, as they are one of key components of the methodology that follows. GP generalizes multivariate Gaussian distribution to a distribution over infinite dimensional vector of functions, such that every finite sample of function values is jointly Gaussian distributed. Due to the Bayesian nature of the solution, GPs are able to consider both epistemic (limited data) and aleatoric (stochasticity of the true model) uncertainty. A GP is fully specified with the prior mean function  $\mu(\cdot)$  and the positive semi-definite kernel function  $k(\cdot, \cdot)$ , and the GP regression aims to model an unknown set of functions  $\mathbf{f} : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_y}$  given some noisy observations  $y_i = f_i(\mathbf{x}) + \epsilon_i$ ,  $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ . For  $\mathbf{f} = [f_1, \dots, f_{n_y}]$ , this could be expressed as:

$$f_i \sim \mathcal{GP}(\mu_i(\cdot), k_i(\cdot, \cdot)), \quad (12)$$

where the prior mean  $\mu_i$  function provides knowledge of the mean of a test point prior to observing data, the kernel function  $k_i$  expresses the covariance between points. Let a number of data points be available, the posterior distribution at a test point  $x^*$  is then found by the conditional distribution given on the available  $N$  noisy data for each output  $i$ :  $\mathcal{D}_i := \{\mathbf{X}, \mathbf{y}_i\}$ , with  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{n_x \times N}$  and  $\mathbf{y}_i = [y_{i1}, \dots, y_{iN}]^T \in \mathbb{R}^N$ . Then, the posterior's mean  $m_i(\mathbf{x}^*)$  and variance  $\Sigma_i(\mathbf{x}^*)$  of the test point  $\mathbf{x}^*$  are

$$\begin{aligned} m_i(\mathbf{x}^*) &= \mathbb{E}(f_i^* | \mathbf{x}^*, \mathcal{D}) = k_i(\mathbf{x}^*, \mathbf{X}) [K_i(\mathbf{X}, \mathbf{X}) + \sigma_i^2 \mathbf{I}]^{-1} (\mathbf{y}_i - \mathbf{1} \mu_i(\mathbf{x}^*)) + \mu_i(\mathbf{x}^*) \\ \Sigma_i(\mathbf{x}^*) &= \mathbb{V}(f_i^* | \mathbf{x}^*, \mathcal{D}) = k_i(\mathbf{x}^*, \mathbf{x}^*) - K_i(\mathbf{x}^*, \mathbf{X}) [K_i(\mathbf{X}, \mathbf{X}) + \sigma_i^2 \mathbf{I}]^{-1} K_i(\mathbf{X}, \mathbf{x}^*), \end{aligned} \quad (13)$$

Notice that  $\Sigma_i$  is the variance of the noise-free prediction, if the noise is taken into consideration then variance should be  $\Sigma_i + \sigma_i^2$ , and  $\sigma_i^2$  is many times treated as a hyperparameter for the GP. The mean and covariance function (kernel) define the prior GP:

$$\mathbf{f}(\mathbf{x}^*) \sim \mathcal{N}(\mathbf{m}(\mathbf{x}^*), \mathbf{\Sigma}(\mathbf{x}^*)), \quad (14)$$

with  $\mathbf{m}(\mathbf{x}^*) = [m_1, \dots, m_{n_y}]$  and  $\mathbf{\Sigma}(\mathbf{x}^*) = \text{diag}([\Sigma_1(\mathbf{x}^*), \dots, \Sigma_{n_y}(\mathbf{x}^*)])$ . Common choices for the kernel  $k_i(\cdot, \cdot)$  is the squared-exponential (SE) covariance function:

$$k_i(\mathbf{x}, \mathbf{x}') = \sigma_i^{\alpha^2} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}')^T \mathbf{\Lambda}_i (\mathbf{x} - \mathbf{x}')\right) \quad (15)$$

and Matérn class of covariance functions, with special cases the Matérn 3/2 and Matérn 5/2:

$$\begin{aligned} k_i^{3/2}(\mathbf{x}, \mathbf{x}') &= \sigma_i^{\alpha^2} \left(1 + \sqrt{3}r_i\right) \exp(-\sqrt{3}r_i) \\ k_i^{5/2}(\mathbf{x}, \mathbf{x}') &= \sigma_i^{\alpha^2} \left(1 + \sqrt{5}r_i + \frac{5}{3}r_i^2\right) \exp(-\sqrt{5}r_i), \end{aligned} \quad (16)$$

with  $r_i = \sqrt{(\mathbf{x} - \mathbf{x}')^T \mathbf{\Lambda}_i (\mathbf{x} - \mathbf{x}')}$ . The parameters  $\sigma_i^{\alpha^2}$ ,  $\mathbf{\Lambda}_i = \text{diag}([\lambda_1, \dots, \lambda_{n_x}])$  and the noise level  $\sigma_i^2$  are hyperparameters for the GP and, to estimate them, a maximum likelihood estimation is carried out. In this work, we wish to estimate an a priori unknown function  $\mathbf{f}$  with GP by collecting measurements during the operation.

### 3.3. Gaussian Process for Model Mismatch

In this work, the constraints are not imposed based on the approximated parametric uncertainty, as it is shown that such an approach would suffer in small data sets. For this reason, the model will be used as a prior for the GP that will approximate the model mismatch. Since the real system states affect the feasible space, here we propose approximating constraints instead of the real system as a whole. Specifically, the GP's prior for

the constraints is computed using (8) first discretizing it, and then constraints can be computed. As a result, the constraints based on the model are:

$$\begin{aligned}\hat{\mathbf{x}}(t_0) &= \mathbf{x}(0) \\ \hat{\mathbf{f}}(\hat{\mathbf{x}}, \hat{\mathbf{x}}, \mathbf{u}, \boldsymbol{\theta}) &= 0 \\ \hat{g}_i(\hat{\mathbf{x}}, \mathbf{u}) &\in \mathbb{X}, \forall i \in \{1, \dots, n_g\},\end{aligned}\tag{17}$$

then with some abuse of the notation the symbol  $\hat{g}_i$  is used for the constraints associated with the available model:

$$\hat{g}_i(\hat{\mathbf{x}}, \mathbf{u}) = \hat{g}_i(\mathbf{u}).\tag{18}$$

When, constraints  $g_i(\mathbf{u})$  are observed, a GP can be built using as prior the predictions of the model  $\hat{g}_i$ . The prior mean  $\mu_i$  of the  $i$ -th constraints is  $\mu_i = \hat{g}_i$  and now the hyperparameters of the GP can be found for this hybrid model.

Given there are  $N$  available experimental points  $\mathcal{D}_i := \{\mathbf{U}, \mathbf{g}_i\}$ , with  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_N] \in \mathbb{R}^{n_u \times N}$  and  $\mathbf{g}_i = [g_i(\mathbf{u}_1), \dots, g_i(\mathbf{u}_N)]^T \in \mathbb{R}^N$ , the mean and variance of the posterior for the (13) can now be written as:

$$\begin{aligned}m_{g_i}(\mathbf{u}^*) &= \mathbb{E}(g_i^* | \mathbf{u}^*, \mathcal{D}_i) = k_i(\mathbf{u}^*, \mathbf{U}) [K_i(\mathbf{u}, \mathbf{U}) + \sigma_i^2]^{-1} (\mathbf{g}_i - \mathbf{1}\hat{g}_i(\mathbf{u}^*)) + \hat{g}_i(\mathbf{u}^*) \\ \Sigma_{g_i}(\mathbf{u}^*) &= \mathbb{V}(g_i^* | \mathbf{u}^*, \mathcal{D}_i) = k_i(\mathbf{u}^*, \mathbf{u}^*) - K_i(\mathbf{u}^*, \mathbf{U}) [K_i(\mathbf{U}, \mathbf{U}) + \sigma^2]^{-1} K_i(\mathbf{U}, \mathbf{u}^*),\end{aligned}\tag{19}$$

Then we wish to satisfy the constraints with a given probability. To satisfy the chance constraints, a distributionally robust reformulation is employed via the Chebyshev-Cantelli theorem [34], (see 3.1) using the mean and variance of the constraints.

**Theorem 3.1.** ([34]) Consider a chance probabilistic constraint of the form

$$\mathbb{P}(q \leq 0) \geq 1 - \epsilon, \quad \epsilon \in (0, 1)\tag{20}$$

where  $q \in \mathbb{R}^{n_q}$  is some random variable. Let  $\mathcal{Q}$  be a family of distributions with mean  $\mu_q$  and variance  $\Sigma_q$ . Then for any  $\epsilon \in (0, 1)$ , the distributionally robust probabilistic constraint

$$\inf_{q \sim \mathcal{Q}} \mathbb{P}(q \leq 0) \geq 1 - \epsilon\tag{21}$$

is equivalent to the following constraint:

$$\mu_q + r\sqrt{\Sigma_q} \leq 0,\tag{22}$$

with  $r = \sqrt{\frac{1 - \epsilon}{\epsilon}}$ .

This theorem provides a conservative estimate for the constraints as it holds for the whole family of  $\mathcal{Q}$ .

**Remark.** It should be noted that this theorem was proposed for general constraints. Additionally the parameter  $r$  is often used as a design parameter to reduce the conservatism of this formulation, however loosing any respected guarantees.

This reformulation relies highly on the accuracy of the GP's mean and variance and the initial small amount of data points may significantly affect the result. Notice that even though the result in Theorem 3.1 is conservative for all the family of distribution with mean  $\mu_q$  and  $\Sigma_q$ , the wrong estimation of mean and variance may lead to infeasible designs.

### 3.4. Trust Regions

In this work, different GPs are utilized to tackle the objective function’s parametric uncertainty and the structural mismatch between the model and process. This section focuses on how the GP can be trained and provide guarantees for the latter case. Mainly, trust regions are utilized to update the GPs for the model-mismatch locally and not globally. First, we introduce the assumption and the bound that it is usually employed, then we show that these bounds may not be used globally using a motivating simple example.

It is common to achieve bounds for the unknown function, using the following assumption:

**Assumption 3.1.** [59, 5] *The unknown function  $f_i$  has bounded norm in reproducing kernel Hilbert space (RKHS)  $\mathcal{H}_k$  [55], induced by a the continuously differentiable kernel  $k_i$ ,  $\|f_i\|_{k_i} \leq B_{f_i}$ ,  $B_{f_i} > 0$ .*

Note that Assumption 3.1 does not make any probabilistic assumption on  $f_i$  but it requires to have a bounded norm in the RKHS or be drawn by a GP prior [57]. Given assumption 3.1, reliable confidence intervals on  $f_i$  can be found:

**Lemma 3.2.** [57] *Let a function  $f_i$  that satisfies Assumption 3.1 with  $k_i$ ,  $\|f_i\|_{k_i} \leq B_{f_i}$  and the measurement noise is a  $\sigma$ -sub-Gaussian noise. Then  $|m_i(\mathbf{x}) - g_i(\mathbf{x})| \leq \beta_i^{(N)} \sigma_i(\mathbf{x})$  with probability higher than  $1 - \delta$ , and  $\beta_i^{(N)} = B_{f_i} + 4\sigma\sqrt{\gamma_N + 1 - \log(\delta)}$ , with  $\gamma_N$  being the information capacity of  $k_i$ .*

The scaling factor  $\beta_i^{(N)}$  in Lemma 3.2 depends on the kernel’s  $k_i$  information capacity  $\gamma_N$ . It is the maximum amount of mutual information [57] that can be achieved using the noisy measurement  $\mathcal{D}_i$  for  $f_i$ . Due to generic nature of the systems, the Assumption 3.1 may not hold globally. For that reason we will aim to use the GP locally instead.

To motivate the need of local approximations and trust regions the following example is presented.

#### 3.4.1. Importance of Trust Regions: A simple motivating example

We first provide a simple motivating example that illustrates the reasons why such approximation is not cautious globally, and then the trust regions for our problem are introduced.

Let a simple function  $f(x) = x \sin(x)$  be assumed as the ‘true’ model, with  $x \in \mathbb{R}$  and the collected data are corrupted by a Gaussian noise  $\epsilon_f$  with standard deviation 0.01,  $y = f(x) + \epsilon_f$ . A GP is used to approximate the data points, specifically a squared-exponential (SE) covariance function is used and its hyperparameters are fitted using maximum likelihood estimation (with multi-starts to avoid local optima). The mean  $m$  and variance  $\Sigma$  of the GP for each  $x$  are illustrated in Fig 2.

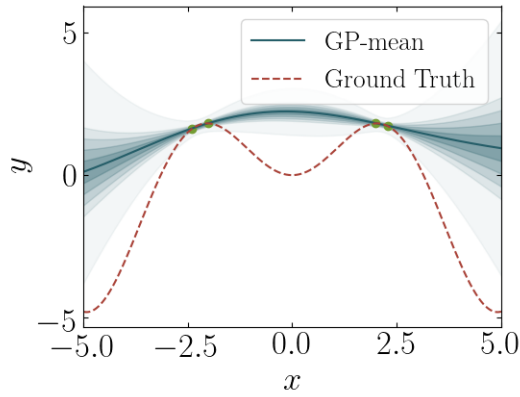


Figure 2: The ground truth (red dashed line) is the function that the GP approximates given the available data. The GP predictions for the mean ( $\mu$ ) is the blue line with variance  $\Sigma$ , and the shaded areas represents the area between  $m \pm \sqrt{\Sigma}$ ,  $m \pm \sqrt{\Sigma}$ ,  $\mu \pm 3\sqrt{\Sigma}$ ,  $m \pm 4\sqrt{\Sigma}$  and  $m \pm 10\sqrt{\Sigma}$ .

The shaded areas show the  $m + r\sqrt{\Sigma}$  for  $r = 1, 2, 3, 4$  and  $10$ . Using the Theorem 3.1 and the Lemma 3.2, we would expect the ‘true’ model (red dashed line) to lie within these shaded areas. Specifically, the Chebyshev-Cantelli theorem states that for  $r = 9.94$ ,  $x$  should lie within the shaded area with probability 99%. However, due to the shape of the ‘true’ model and the collected data, such guarantee is lost as the predicted variance and mean are poorly computed. This motivates the further restriction of the proposed optimization problem, as it is clear that the GP cannot be used globally with small data-sets and it is a paramount importance to satisfy the constraints with high probability.

#### 3.4.2. Trust region updates

The use of local models has extensively been employed in the literature of derivative-free optimization [13, 12], where trust regions (TRs) are used to bound the predictions of the surrogate function of the respected objective function around the current point. The size of the TR is updated in every iteration according to a ratio of the objective’s actual reduction, over the predicted reduction. In this work the same philosophy is used for the approximation of the constraints: A TR is defined for each constraint in the problem and it is updated according to the accuracy of the predictions on the new observations. The TR is defined as:

$$(\mathbf{u} - \mathbf{u}_k)^T (\mathbf{u} - \mathbf{u}_k) \leq \mathcal{R}_i, \quad \forall i \in \{1, \dots, n_g\} \quad (23)$$

where  $\mathbf{u}_k$  is the operating point (experimental condition) and  $\mathcal{R}_i$  the radius of the trust region. The rule for updating the TR is described in Algorithm 1.

---

**Algorithm 1** Trust region updates

---

**Input:** Give parameter  $\eta_1, \eta_2, t_2 < 1 < t_1$  and  $R_i$ . New observation  $g_i(\mathbf{u}^*)$  and model prediction  $\hat{g}_i(\mathbf{u}^*)$  for  $\mathbf{u}^*$ .

1. Compute  $\rho_i = \left( \frac{(g_i(\mathbf{u}^*) - \hat{g}_i(\mathbf{u}^*))}{g_i(\mathbf{u}^*)} \right)^2$
2. If  $\rho_i \leq \eta_1$ :  $\mathcal{R}_i = t_1 \mathcal{R}_i$
3. If  $\rho_i \geq \eta_2$ :  $\mathcal{R}_i = t_2 \mathcal{R}_i$
4. If  $\eta_1 \leq \rho_i \leq \eta_2$ :  $\mathcal{R}_i = \mathcal{R}_i$

**Output:** New radius  $\mathcal{R}_i$ .

---

According to Algorithm 1, the radius  $\mathcal{R}_i$  is updated according to the accuracy ( $\rho_i$ ) of the new predicted point. If the percentage error is high, i.e.  $\rho_i \leq \eta_1$  then the trust region is reduced if the percentage error is low then the trust region is increased.

The values of the parameters  $\eta_1, \eta_2, t_2 < 1 < t_1$ , can be seen as notion of conservatism for the optimization. If the parameters  $t_1$  and  $t_2$  are closer to 1, then the scheme becomes more conservative, as the trust region changes very little at each iteration. The values for  $\eta_1$  and  $\eta_2$  are the bounds that we want for the approximation of the physical system. If their values are selected to be large then relaxation is imposed to the scheme, as the radius will increase even if the GP is not good enough.

The predictions are now inside the trust region and the computed mean and variance are accurate locally. This way the Lemma 3.2 and Theorem 3.1 may be applied to guarantee the feasibility of the experimental design.

### 3.5. GP-based Uncertainty Propagation for the Information Metric

The GPs are further utilized to compute the metric of information/variance accounting for the parametric uncertainty. GP will be used to directly approximate the metric as function of the optimization variables and the model parameters. Then, model parameter marginalization is used to compute the expected value and variance of the information metric with respect to a computed uncertainty of the parameters. The use of surrogates to propagate the uncertainty is not new in MBDoE, where polynomial chaos expansion is mainly used. Similar approach was followed in [36], but for the case of model discrimination.

Consider the objective function of the experimental design at the  $k^{th}$  experiment:

$$\begin{aligned}
 J(\mathbf{u}, \boldsymbol{\theta}) &= \psi \left( \frac{\partial \hat{\mathbf{x}}(t_N)}{\partial \boldsymbol{\theta}}^\top \Sigma_{exp} \frac{\partial \hat{\mathbf{x}}(t_N)}{\partial \boldsymbol{\theta}} + \mathbf{M}_0 \right) \\
 & \quad s.t. \\
 & \quad \hat{\mathbf{x}}(t_0) = \mathbf{x}(0) \\
 & \quad \hat{\mathbf{f}}(\dot{\hat{\mathbf{x}}}, \hat{\mathbf{x}}, \mathbf{u}, \boldsymbol{\theta}) = 0 \\
 & \quad \hat{\mathbf{y}} = \hat{\mathbf{h}}(\hat{\mathbf{x}}, \mathbf{u})
 \end{aligned} \tag{24}$$

Notice that in this case the training set for the GP is generated in-silico using the available approximated model, hence a sufficiently large data-set can be generated [36]. This is a main difference in the training of this GP with the ones used for the model mismatch, as there is no need for local approximations. Given the

uncertainty region that has been computed (see section 2.1) the uncertainty can not be directly propagated via the GP. The posterior distribution will not be Gaussian and a Monte-Carlo sampling is needed. For that reason further approximations are utilized; a very common approximation is to use Taylor expansion (see Appendix A.1 with an additional example), where the expected value and variance of the objective is computed with respect to the uncertain parameters  $\theta$ .

The expected value of the objective in MBD<sub>oE</sub> (the information metric) can be substituted by the mean function of the GP. Note that the objective function can also incorporate a penalty term for large variations [8]. The objective function is now reformulated as:

$$\hat{m}_J(\mathbf{u}, \boldsymbol{\mu}_{\hat{\theta}}, \boldsymbol{\Sigma}_{\hat{\theta}}) + \alpha_J \sqrt{\hat{\Sigma}_J(\mathbf{u}, \boldsymbol{\mu}_{\hat{\theta}}, \boldsymbol{\Sigma}_{\hat{\theta}})}, \quad (25)$$

where  $\hat{m}_J(\mathbf{u}, \boldsymbol{\mu}_{\hat{\theta}}, \boldsymbol{\Sigma}_{\hat{\theta}})$  and  $\hat{\Sigma}_J(\mathbf{u}, \boldsymbol{\mu}_{\hat{\theta}}, \boldsymbol{\Sigma}_{\hat{\theta}})$  are the mean and variance of the GP after the propagation of the uncertainty of the parameters  $\theta$ ,  $\alpha_J$  is a constant variable. The value of  $\hat{m}_J(\mathbf{u}, \boldsymbol{\mu}_{\hat{\theta}}, \boldsymbol{\Sigma}_{\hat{\theta}})$  represents the information metric in expectation when uncertain parameters have mean  $\boldsymbol{\mu}_{\hat{\theta}}$  and  $\boldsymbol{\Sigma}_{\hat{\theta}}$ .

### 3.6. Approximate Optimal Model-Based Design of Experiments

The optimization problem that is solved to find the new design  $\mathbf{u}^*$  given the previous design  $\mathbf{u}_k$  and the parameter's computed mean ( $\boldsymbol{\mu}_{\hat{\theta}}$ ) and variance ( $\boldsymbol{\Sigma}_{\hat{\theta}}$ ) is:

$$\begin{aligned} \min_{\mathbf{u}} \quad & \hat{m}_J(\mathbf{u}, \boldsymbol{\mu}_{\hat{\theta}}, \boldsymbol{\Sigma}_{\hat{\theta}}) - \alpha_J \sqrt{\hat{\Sigma}_J(\mathbf{u}, \boldsymbol{\mu}_{\hat{\theta}}, \boldsymbol{\Sigma}_{\hat{\theta}})} \\ \text{s.t.} \quad & \\ & m_{g_i}(\mathbf{u}) + r \sqrt{\Sigma_{g_i}(\mathbf{u})} \leq 0 \\ & (\mathbf{u} - \mathbf{u}_k)^T (\mathbf{u} - \mathbf{u}_k) \leq \min(\mathcal{R}_i) \\ & \forall i \in \{1, \dots, n_g\}, \end{aligned} \quad (26)$$

where  $3 > \alpha_J > 0$  is used to add exploration to the design. This parameter can be seen as the upper confidence bound formulation in Bayesian optimization [18]. The larger the value of  $\alpha_J$  the more significant the exploration will be. In this work, the optimization will mainly focus on exploitation, and for that reason  $\alpha_J$  will have the value of 0.1. If the user wants to emphasize to exploration then values between 2 and 3 should be used, inspired by the upper confidence bound methods [27].

The Algorithm 2 summarizes the proposed framework.



**Algorithm 2** Approximate optimal model-based design of experiments

**Input:** Give parameter  $\eta_1, \eta_2, t_2 < 1 < t_1, \alpha_J$  and  $tol_1$ . For  $i^{th}$  constraint give  $N$  data  $\mathcal{D}_i := \{\mathbf{U}, \mathbf{g}_i\}$ , with  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_N] \in \mathbb{R}^{n_u \times N}$  and  $\mathbf{g}_i = [g_i(\mathbf{u}_1), \dots, g_i(\mathbf{u}_N)]^T \in \mathbb{R}^N$  and radius  $R_i$ . Posterior distribution for the model parameters  $\Theta$

1. Compute the model predictions for each constraint,  $\hat{g}_i$  for all  $\mathbf{U}$ .
2. Train GP for the constraints using  $\hat{g}_i(\mathbf{u})$  as the prior.
3. Generate in-silico data for the objective function:  $J : \mathbf{U} \times \Theta \rightarrow \mathbb{R}$
4. Train GP to approximate  $J$  and propagate uncertainty using equations (A.2-A.5) & (25)
5. Solve Optimization (26) that results in  $\mathbf{u}^*$

**Output:** Get new design  $\mathbf{u}_{k+1} = \mathbf{u}^*$ .

The steps that followed to solve the MBDoE problem using GPs are given below: First in **step 1** and **step 2** a Gaussian process is trained to approximate the constraints for the physical system. Here experimental data gathered from the process are used. Notice that first the model predictions for the model are computed as the model is used as prior for the GP (see Section 3.3). Then, in-silico artificial data are generated in **step 3** to approximate the objective function using a GP in **step 4** (see Section 3.5 and Appendix A.1). In **step 4**, the uncertainty is propagated (see Section 3.5 and Appendix A.1). Finally the optimization is solved in **step 5**, using (26). In this step, the trust regions are applied, in order to restrict the feasible space and maintain a good local approximation.

The proposed framework, GP-MBDoE, first shown in Fig 1b can now be described in Algorithm 3. Algorithm 3 represents the proposed framework, whereas the Algorithm 2 is utilized at the step of that the next optimal experimental design.

**Algorithm 3** Proposed Framework

**Input:** Give parameter  $\eta_1, \eta_2, t_2 < 1 < t_1, \alpha_J$  and  $t_3 < 1$ . For  $i^{th}$  constraint give  $N$  data  $\mathcal{D}_i := \{\mathbf{X}, \mathbf{y}_i\}$ , with  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_N] \in \mathbb{R}^{n_u \times N}$  and measurements  $\mathbf{g}_i = [g_i(\mathbf{u}_1), \dots, g_i(\mathbf{u}_N)]^T \in \mathbb{R}^N$  for each constrained  $i$ , and radius  $\mathcal{R}_i$ .

Select design  $\mathbf{u}_0$

**for**  $k = \{1, \dots\}$ :

1. Perform Parameter estimation to obtain  $\hat{\boldsymbol{\theta}}$ .
2. Approximate posterior distribution  $\Theta$  and compute mean  $\boldsymbol{\mu}_{\hat{\boldsymbol{\theta}}}$  and variance  $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$ .
3. **if**  $\|\mathbf{u}_{k+1} - \mathbf{u}_k\| \leq tol_1$  **or** statistics are satisfied:  
Exit
4. Find the new design  $\mathbf{u}_{k+1}$  using **Algorithm 2** with inputs  $(\mathbf{U}, \mathbf{g}_i, \boldsymbol{\mu}_{\hat{\boldsymbol{\theta}}}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}, \mathcal{R}_i, \alpha_J)$ .
  - (i) Update  $\mathcal{R}_i$  using **Algorithm 1** with inputs  $(t_1, t_2, \eta_1, \eta_2, \mathcal{R}_i)$ .
  - (ii) **if**  $g_i(\mathbf{u}_k) > 0$ :  $\mathcal{R}_i = t_3 \mathcal{R}_i$  and  $\mathbf{u}_{k+1} = \mathbf{u}_k$
5. Perform new experiment and update  $\mathbf{U}, \mathbf{g}_i$  and  $\mathbf{y}_j$ .

**Output:** Experiments finished.

The steps followed in this work as illustrated in Fig. 1b are the following for each iteration  $k$ : Given the preliminary data (and the updated data), **Parameter Estimation** is performed, specifically **Step 1** and **Step 2** are used to obtain the best parameters  $\boldsymbol{\theta}$  given the initial data points we have and also approximate the posterior distribution. As it is described in section 2.1, the parametric uncertainty that is obtained is not suitable for the constraint satisfaction. However, the parametric uncertainty can be used for the objective of the experimental design. If the **Termination criteria** are satisfied (see Section 2.4) then in **Step 3** the experimental procedure stops. Algorithm 2 is then performed in **Step 4**, where the new optimal design is found. This step also entails the construction of the GP for the model mismatch using the parametric model as prior (**Approximate the Model Mismatch** in Fig. 1b), the GP-based uncertainty propagation for the objective function (**GP-based Uncertainty Propagation** in Fig. 1b) and solution of the optimization for the computation of the new design (26) (**Solve the Optimization Problem** in Fig. 1b). In **Step 4.i-ii** the **Update Trust region** (Fig. 1b) step is performed following Algorithm 1. After that in **Step 5** the new experiment is performed and the last step contains a backtracking that is used when an infeasible design happen where the next design is return to the previous value and the trust region of that constraint shrinks.

#### 4. Case Studies

The proposed methodology (GP-MBDoE) is illustrated using two case studies with different scenarios simulated in silico:

- Case Study 1: a plug flow reactor where that the assumed mechanism is different than the ‘real’ one with an additional constant disturbance. The model contains 4 parameters and has 2 decision variables.

- Case Study 2: a plug flow reactor that a Nucleophilic aromatic (SnAr) substitution [26] reaction takes place, where the concentrations have a non-constant disturbance affect by the inlet concentrations. Such phenomenon can be observed in case a pump malfunction occurs. The model contains 8 parameters and has 4 decision variables.

For each case study, two models structures are postulated: (i) a structure to simulate the physical system, which is employed to perform the in silico experiments and (ii) a misspecified model structure with corrupted measurements. Once a preliminary instance of the model parameters for each of the structurally incorrect models is obtained, both case studies involve an experimental design stage. In order to evaluate our methodology, we compare the GP-MBDoE with the standard methods that consider uncertainty:

- Monte Carlo-based robust MBDoE (MC-MBDoE) under parametric uncertainty, where the Monte Carlo using sampling from the posterior distribution of the parameters to calculate the backoffs [19]. A sampling of the expected uncertainty domain of the parameters needs to be performed to compute the effect of the possible realizations of the unknown parametric set on the model's state variables. Stochastic simulations are performed to evaluate the uncertain parameter's effect and compute the tightening of the constraints (backoffs). The optimization is solved, and the percentage of the constraint violation is computed, and the backoffs are re-estimated, and the stochastic simulation is performed. This process continues until the backoffs have converged. Notice that other more sample efficient methods could be used [8], but would require additional assumption for the distribution of the constraints.
- Disturbance estimation-based robust MBDoE (DE-MBDoE), where the structural model mismatch will be modeled by a disturbance, a method that is originated by the MPC literature [38] and applied to online MBDoE [20]. This method can be seen as a simplification to our approach where instead of using a GP, a constant disturbance is used and updated at each step. Estimation of a constant disturbance is performed using the difference between the measurement and predictions.

The GP-MBDoE required to construct different GPs for the objective function and the constraints. The GP for objective function requires simulations of the approximated model. For that reason, a sufficiently high number of data points should be used. The kernel used for the objective function is the Matérn 5/2 class of covariance function with 200 and 400 points for each case study, respectively. For the computational times see (Section 5). For the approximation of the model mismatch, where only a limited number of data points can be used, the simple SE covariance function is implemented.

#### 4.1. Case Study 1: Flow Reactor: 'Wrong Mechanism'

A flow reactor is assumed where the system can be described by the following set of DAEs:

$$-\frac{F}{A} \frac{dc_i}{dz} + \sum_{j=1}^{N_r} a_{ij} r_j = 0, \quad (27)$$

where  $c_i$  is the concentration (mol/L) of the  $i^{th}$  reagent,  $a_{ij}$  is the stoichiometric coefficient of the  $i^{th}$  reagent in the  $j^{th}$  reaction with  $r_j$  reaction rate. The length of the reactor is  $L = 25\text{cm}$  with surface  $A = 1.2\text{cm}^2$  and  $F(\text{mL}/\text{min})$  is the inlet flowrate  $F$ . The temperature dependency of the reaction rate is described using Arrhenius law:

$$k_j^r = A_j \exp\left(\frac{-E_{aj}}{R} \frac{1}{T + 273.15}\right), \quad (28)$$

where  $A_j$  denotes the frequency factor and  $E_{aj}$  the activation energy. In this setting the design variables are the inlet flowrate  $F$  and temperature  $T$ , i.e.  $\mathbf{u} = [F, T]$ . The formulation in (28) tends to be sloppy [11], for that reason the following formulation is used instead:

$$k_j^r = \exp\left(k_j^0 - \frac{E_j 10^4}{R} \left(\frac{1}{T + 273.15} - \frac{1}{90 + 273.15}\right)\right), \quad (29)$$

The mechanism that takes place in the true model is  $\left\{ \begin{array}{l} A \xrightarrow{k_1} B \xrightarrow{k_2} C \\ A \xrightarrow{k_3} C \end{array} \right\}$ , there is an additional constant disturbance  $w$  of 0.1 M added to the concentrations. The model that is assumed in the design has the following mechanism  $\left\{ A \xrightarrow{k_1} B \xrightarrow{k_2} C \right\}$ , where one reaction is ignored ( $\left\{ A \xrightarrow{k_3} C \right\}$ ). We assume to have upper and lower bounds on concentration measurements of  $c_C$ , i.e.  $g_1 = c_C - 0.4 \leq 0$  mol/L and  $g_2 = c_C - 0.1 \geq 0$  mol/L to be satisfied, satisfied with probability 90%, respectively. The measurements are corrupted by normally distributed noise with standard deviation  $[\sigma_A, \sigma_B, \sigma_C] = [0.039, 0.14, 0.05]$ . Additionally, reaction rate  $i$  has the following form:

$$k_j^r = k_j^0 \exp\left(\frac{-E_{aj}}{R} \frac{1}{T + 273.15}\right) \quad (30)$$

The parameters for the true model are depicted in Table 1:

Table 1: Parameter values assumed for the true kinetic model

<i>Parameter</i>	Value
$k_1^0$ (min <sup>-1</sup> )	8
$E_{a1}$ (kJ/mol)	29
$k_2^0$ (min <sup>-1</sup> )	5
$E_{a2}$ (kJ/mol)	35
$k_3^0$ (min <sup>-1</sup> )	3
$E_{a3}$ (kJ/mol)	32

The concentration of the the inlet reagent A is constant at 2 mol/L. The experimental design variables  $T, F$  are bounded with  $T(^{\circ}C) \in [60, 100]$  and  $F(\text{mL}/\text{min}) \in [0.004, 0.008]$ . Latin hypercube is used to generate an initial set of 5 experiments (Table 2).

Table 2: Initial set of experimental conditions for case study 1

$T(^{\circ}C)$	$F(\text{mL}/\text{min})$
79.6	0.0069
97.6	0.0055
62.8	0.0077
88.0	0.0048
69.0	0.0062

The following configurations are discussed:

- MBDoE under parametric uncertainty (MC-MBDoE).

- MBDoE with disturbance estimation (DE-MBDoE)
- GP-based MBDoE (GP-MBDoE)

A D-optimal design criterion is used in all the proposed experiment design configurations. The results follow in the next paragraphs of this section regarding manipulated inputs, simulated profiles and a posteriori statistics on the final parameter estimation. To fairly evaluate all the methodologies, the same method for propagating uncertainty in the objective function is used. Specifically, the posterior distribution of the parameters is approximated with a normal distribution, updated in every iteration. The parameters needed for the Algorithm 3 are depicted in Table 3:

Table 3: Parameters for Algorithm 3

<i>Parameter</i>	Value
$\alpha_J$	0.5
$\eta_1$	$10^{-3}$
$\eta_2$	$10^{-2}$
$t_2$	0.5
$t_1$	2
$t_3$	0.5
$\mathcal{R}_i$	0.3

Typical values for  $t_i$  and  $\mathcal{R}_i$  can be found in [13]. Notice that the trust region is defined for the normalized design variables that are inside  $[-1, 1]$ . All the methods are applied in the same closed loop manner, and the results are presented next. Fig. 3 depicts the constraint satisfaction of the first constraint  $g_1$  for the GP-MBDoE, it is evident that the constraint remain feasible in the all the sequence of experiments even though it approaches the bound of the constraint. The blue shaded area represents the 99.7% confidence intervals of the GP. As expected, due to the design space that is used in every consecutive optimization, the measurements of the true model lie within the confidence interval. Notice that the confidence interval shrinks as more measurements are acquired.

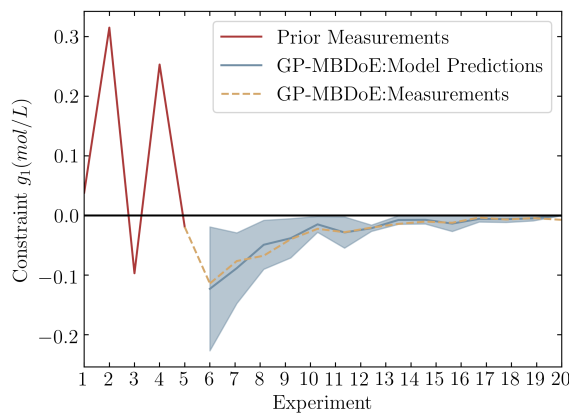
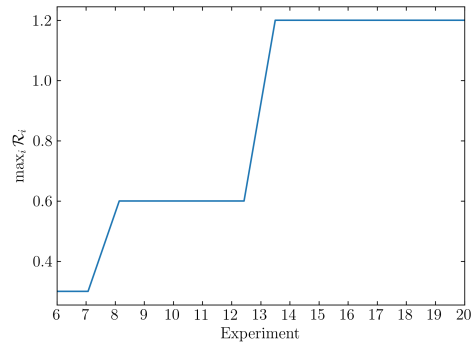


Figure 3: Case study 1: Performance of GP-MBDoE. Blue shaded area represents the 99.7% confidence interval of the GP. The dashed yellow line is the behaviour of the true model and the red lines represent the value for the constraint at the initial experiments.

This behaviour can also be seen in Fig. 4 , where the radius of trust region  $\mathcal{R}_1$  increases as new measurements

Figure 4: Trust region for the constraint  $g_1$ 

are acquired and the GP prediction improves.

On the other hand none of MC-MBDoE and DE-MBDoE formulations managed achieve the desired goal; i.e. remain in the feasible area of the constraints. This is not a surprise as none of the aforementioned method accounted the structural model mismatch properly. It is noticeable from Fig. 5 that the resulted solution from the optimization is predicted to remain feasible with relatively large confidence intervals. However, the MC-MBDoE can account only for a parametric uncertainty, in which the posterior of the parameters has been approximated, and the DE-MBDoE allows the structural model mismatch to be constant that is updated at each iteration. Additionally, in this work Monte-Carlo method is used to solve the stochastic MBDoE problem, which is computationally intensive as in [19]; however approximated solutions can be utilized, e.g. polynomial chaos expansions [34].

Notice that the design with GP-MBDoE defines a test that is now both safe (feasible) and informative for the approximated model as it can be seen in Table 4 that the t-values for the model parameters are larger than the  $t_{ref}$ . The t values and estimated parameters for MC-MBDoE and DE-MBDoE can also be found in Table 5 and Table 6 accordingly.

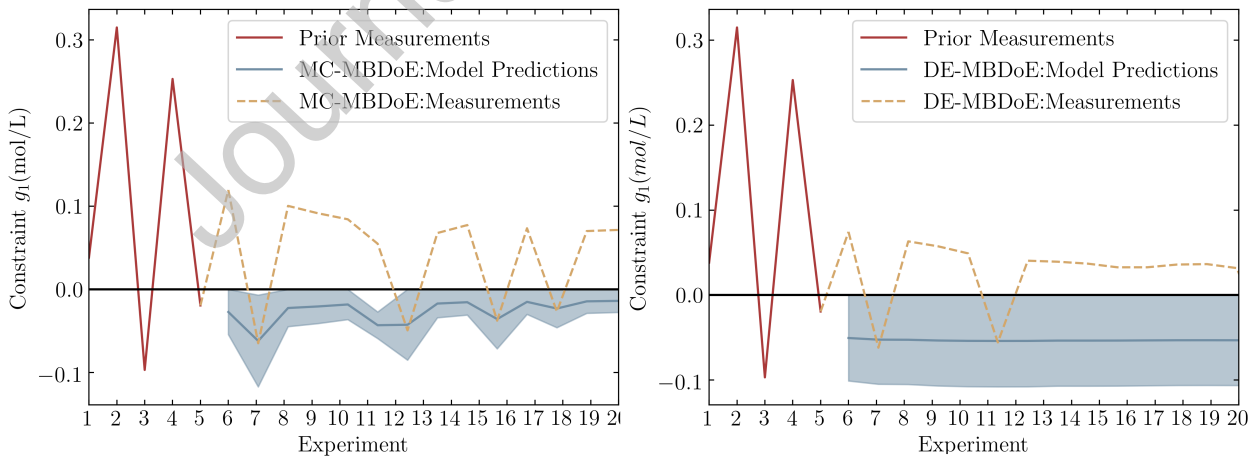


Figure 5: Case study 1: Performance of MC-MBDoE and DE-MBDoE. Blue shaded area represents the computed backoff for each case. The dashed yellow line is the behaviour of the true model and the red lines represent the value for the constraint at the initial experiments.

Table 4: Case study 1: GP-MBDoE. Parameter estimation results in terms of estimated values, 95% confidence intervals and t-values.

Model Parameters	Estimated Values	95% t (final)	$t_{ref}$ (final)	95% t (initial)	$t_{ref}$ (initial)
$k_1^0(\text{min}^{-1})$	6.51	2.38	1.67	1.43*	1.80
$E_{a1}(\text{kJ/mol})$	28.3	37.28	1.67	13.04	1.80
$k_2^0(\text{min}^{-1})$	0.02	6.07	1.67	1.15*	1.80
$E_{a2}(\text{kJ/mol})$	14.4	7.22	1.67	1.93	1.80

Table 5: Case study 1: MC-MBDoE. Parameter estimation results in terms of estimated values, 95% confidence intervals and t-values.

Model Parameters	Estimated Values	95% t (final)	$t_{ref}$ (final)	95% t (initial)	$t_{ref}$ (initial)
$k_1^0(\text{min}^{-1})$	6.48	6.01	1.67	1.43*	1.80
$E_{a1}(\text{kJ/mol})$	2.83	22.01	1.67	13.04	1.80
$k_2^0(\text{min}^{-1})$	0.02	3.09	1.67	1.15*	1.80
$E_{a2}(\text{kJ/mol})$	15.4	7.06	1.67	1.93	1.80

Table 6: Case study 1: DE-MBDoE. Parameter estimation results in terms of estimated values, 95% confidence intervals and t-values.

Model Parameters	Estimated Values	95% t (final)	$t_{ref}$ (final)	95% t (initial)	$t_{ref}$ (initial)
$k_1^0(\text{min}^{-1})$	6.88	3.86	1.67	1.43*	1.80
$E_{a1}(\text{kJ/mol})$	28.4	19.45	1.67	13.04	1.80
$k_2^0(\text{min}^{-1})$	0.01	3.17	1.67	1.15*	1.80
$E_{a2}(\text{kJ/mol})$	9.7	10.41	1.67	1.93	1.80

The parameter are updated in each additive sample that it is obtained. Fig. 6 presents the trajectory in time for the estimated parameter  $E_{a2}$ . The behaviour of the parameters is interesting, as their estimation appear to have similar behaviour with the use of MC-MBDoE, which appears to be the least conservative/restrictive (with the most infeasible designs). Notice that due to the model mismatch there are no ‘true’ parameters for the approximated model, that matches the behaviour of the true model.

Table 7: Case study 1: results in terms of chi-square statistics obtained from the different MBDoE configurations. 95%  $\chi^2$ ,  $\chi_{ref}^2 = 74.47$ .

Method	95% $\chi^2$ ( $\chi_{ref}^2 = 74.47$ )
GP-MBDoE	2018
MC-MBDoE	1965
DE-MBDoE	1888

Table 7 shows that all three methods (i.e. GP-MBDoE, MC-MBDoE, DE-MBDoE) perform similarly in terms of fitting performance. Nevertheless, the GP-MBDoE managed to remain feasible in all the experiments.

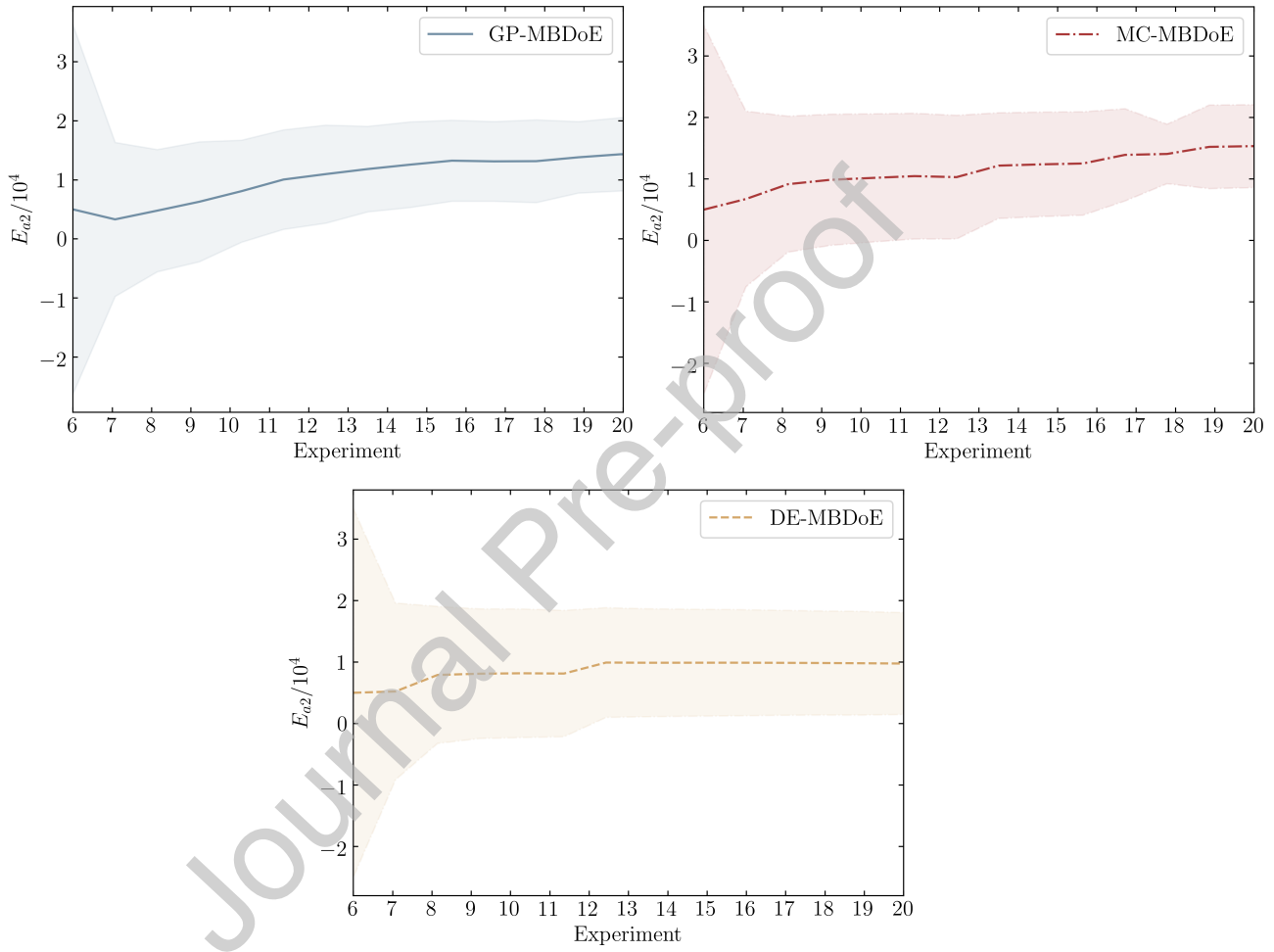


Figure 6: Case Study 1: estimation profiles of  $E_{a2}$  for the different designs. The shaded uncertainty area is represented by the square root of the 95% confidence intervals.



Interestingly the DE-MBDoE performs better with regards of the  $\chi^2$  metric, however the experimental conditions do vary as much as the MC-MBDoE and the GP-MBDoE.

For completeness purposes, we have added the design variables for all the methods (GP-MBDoE, MC-MBDoE, DE-MBDoE) in Appendix C.

#### 4.2. Case Study 2: Flow Reactor: Corrupted Measurements

To further validate the proposed method, a different case study is presented in the following section, where the model will now be assumed to be correct but the disturbance will depend on the experimental conditions. This second case study considers a nucleophilic aromatic substitution (S<sub>N</sub>Ar) of 2,4-difluoronitrobenzene {1} with pyrrolidine {2} in ethanol (EtOH) to give a mixture of the desired product ortho-substituted {3}, para-substituted {4} and bis-product {5} as side products. We chose this reaction as a benchmark to evaluate new systems because it produces 3 different major products that are relevant for pharmaceuticals and fine chemicals [10]. The inlet flowrates of the reagents 2,4-difluoronitrobenzene {1} with pyrrolidine {2} ( $F_1$  and  $F_2$ ), solvent ethanol ( $F_3$ ) and the temperature  $T$  are the design variables:  $\mathbf{u} = [T, F_1, F_2, F_3]$  This case study has been adopted from [26] and the schematic of the reaction is depicted in Fig. 7. The molar balance of the reactor,

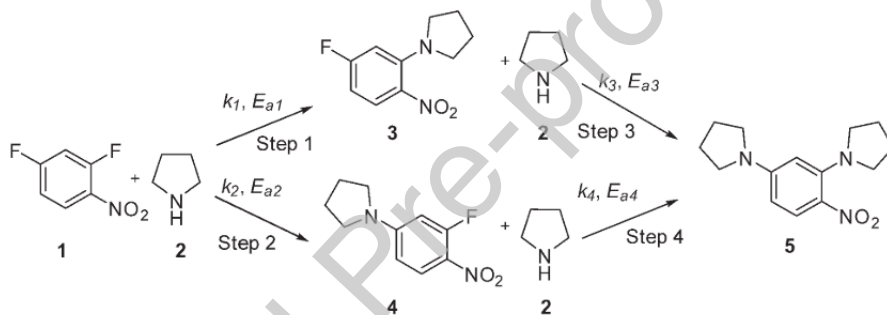


Figure 7: Schematic of the S<sub>N</sub>Ar reaction

using the chemistry from Fig. 7, can be written as:

$$\frac{dc_1}{d\tau} = -r_1 - r_2, \quad \frac{dc_2}{d\tau} = -r_1 - r_2 - r_3 - r_4 \quad (31)$$

$$\frac{dc_3}{d\tau} = r_1 + r_3, \quad \frac{dc_4}{d\tau} = r_2 - r_4, \quad \frac{dc_5}{d\tau} = r_3 + r_4$$

with  $\tau$  (min) being the residence time,  $c_i$  (mol/L) are the molar concentrations for the 2,4-difluoronitrobenzene {1}, pyrrolidine {2}, ortho-substituted {3}, para-substituted {4} and bis-product {5}. The reaction rates follow a power law, and the kinetic constant is parametrized as in Case study 1 (29). The values for the kinetic parameters of the true model are adopted from [26].

In this case study the assumptions of the kinetic models are correct, however there is a malfunction of an equipment (e.g. HPLC, pumps) that results in a disturbance for the concentration of 2,4-difluoronitrobenzene {1}. This systematic error is assumed to be unknown and leads to an overestimation of concentration measurements of 10% for the 2,4-difluoronitrobenzene {1}. An additional concentration measurement noise is considered with zero mean and  $10^{-3}$  mol/L standard deviation for all components. The constraint that is and considered is  $g_1 = c_1 - 0.1 \leq 0.0$  with probability  $\mathbb{P}(g_1 \leq 0) \geq 0.9$ .

Lower/upper bounds on design variables, defining the design space, are reported in Table 8.

Table 8: Case study 2: definition of design space.

Design Variable	Min	Max
$F_1$ (mL/min)	0.3	3.0
$F_2$ (mL/min)	0.3	3.0
$F_3$ (mL/min)	0.3	3.0
$T$ ( $^{\circ}C$ )	60.0	120.

As in the previous Latin hypercube is used to generate the experimental conditions for 5 experiments (Table 9).

Table 9: Case study 2: 5 initial set of experimental conditions for case study 2

	1	2	3	4	5
$T$ ( $^{\circ}C$ )	104.8	129.9	61.1	106.9	83.5
$F_1$ (mL/min)	1.0	2.5	3.0	2.4	1.7
$F_2$ (mL/min)	1.9	2.5	0.9	1.2	1.6
$F_3$ (mL/min)	2.1	1.9	2.7	1.3	1.1

In this case study a D-optimal design is used in all the proposed experimental design configurations. The same parameters for the algorithm used as in case study 1 (Table 3) have been employed. The results follows in the next paragraphs of this section in terms of manipulated inputs, simulated profiles and a posteriori statistics on the final parameter estimation. As in case study 1, all the methodologies are compared with the same method for propagation of uncertainty in the objective function. Additionally, the posterior distribution of the parameters is approximated with a normal distribution, which is updated in every iteration.

All the configurations (MC-MBDoE, DE-MBDoE, GP-MBDoE) are tested in the same online set-up, and the results are presented next. Fig. 8 depicts the constraint satisfaction of the first constraint  $g_1$  for GP-MBDoE. It is evident that the constraint remain feasible in the all the sequence of experiments even though it approaches the bound of the constraint. The blue shaded area represents the 99.7% confidence intervals of the GP. As expected due to the design space that is used in every consecutive optimization the measurements of the true model lie within the confidence interval of predictions of GP. Notice that the confidence interval of the predictions of the concentration shrinks as more measurements are acquired.

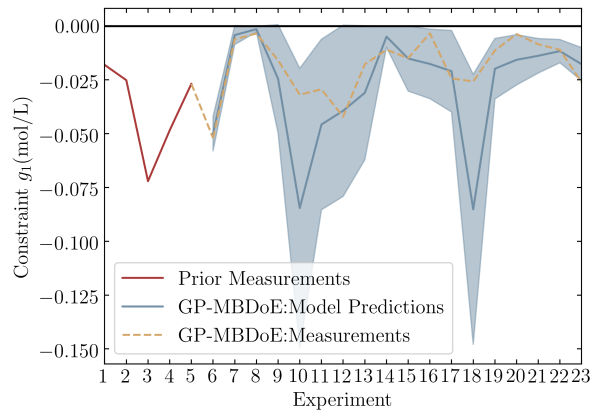


Figure 8: Case study 2: Performance of GP-MBDoE. Blue shaded area represents the 99.7% confidence interval of the GP. The dashed yellow line is the behaviour of the true model and the red lines represent the value for the constraint at the initial experiments.

This behaviour can also be seen in Fig.9, where the trust region radius  $\mathcal{R}_1$  increases as more measurements are acquired.

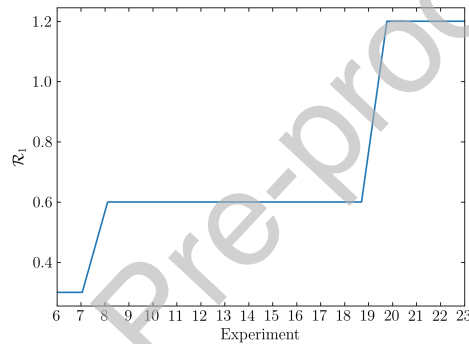


Figure 9: Trust region for the constraint  $g_1$

None of the other formulations managed to remain in the feasible area of the constraint  $g_1$ , as it is depicted in Fig. 10. This is not a surprise as none of the aforementioned methods accounted for the structural model mismatch properly. It is noticeable from Fig. 10 that the resulted solution from the optimization is predicted to remain feasible. However, the MC-MBDoE can account only for parametric uncertainty. The posterior of the parameters has been approximated, and the DE-MBDoE can only allow the structural model mismatch to be constant. Additionally, the solution of the MC-MBDoE is also computationally intensive as it requires the use of Monte Carlo sampling from the confidence region of the parameters; in this case, 1000 samples were sampled at each iteration (more details in [19]).

Notice that the design with GP-MBDoE defines a test that is now both safe (feasible) and informative for the approximated model, as it can be seen in Table 10 that the t-values for the model parameters are larger than the  $t_{ref}$ , which means that they are precisely estimated (very low uncertainty). Nevertheless, high precision does not mean high accuracy, especially in cases with structural mismatch. The t values and estimated parameters for MC-MBDoE and DE-MBDoE can also be found in Table 11 and Table 12 accordingly.

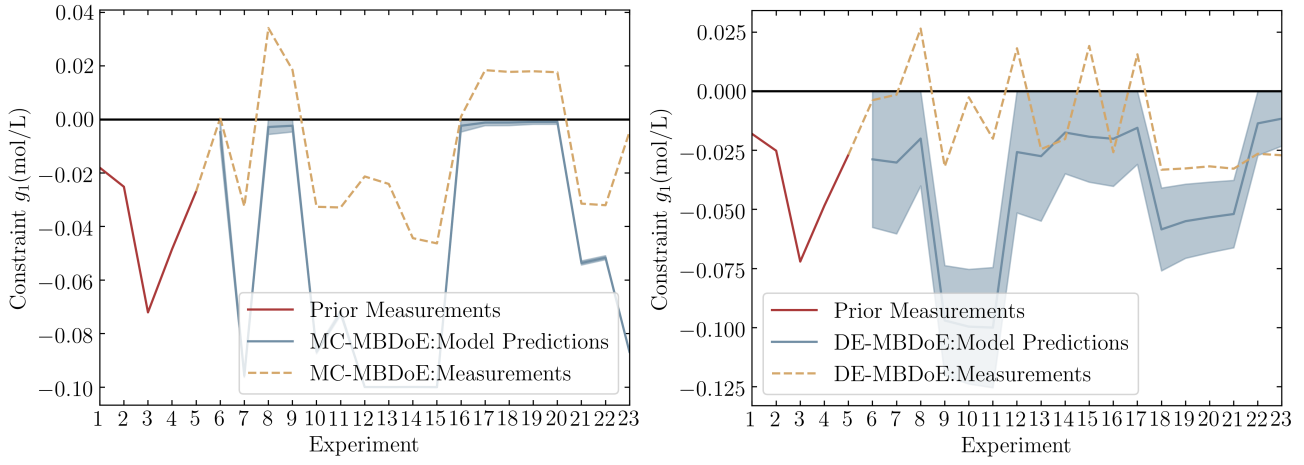


Figure 10: Case study 2: Performance of MC-MBDoE and DE-MBDoE. Blue shaded area represents the computed backoff for each case. The dashed yellow line is the behaviour of the true model and the red lines represent the value for the constraint at the initial experiments.

Table 10: Case study 2, GP-MBDoE. Parameter estimation results in terms of estimated values, 95% confidence intervals and t-values.

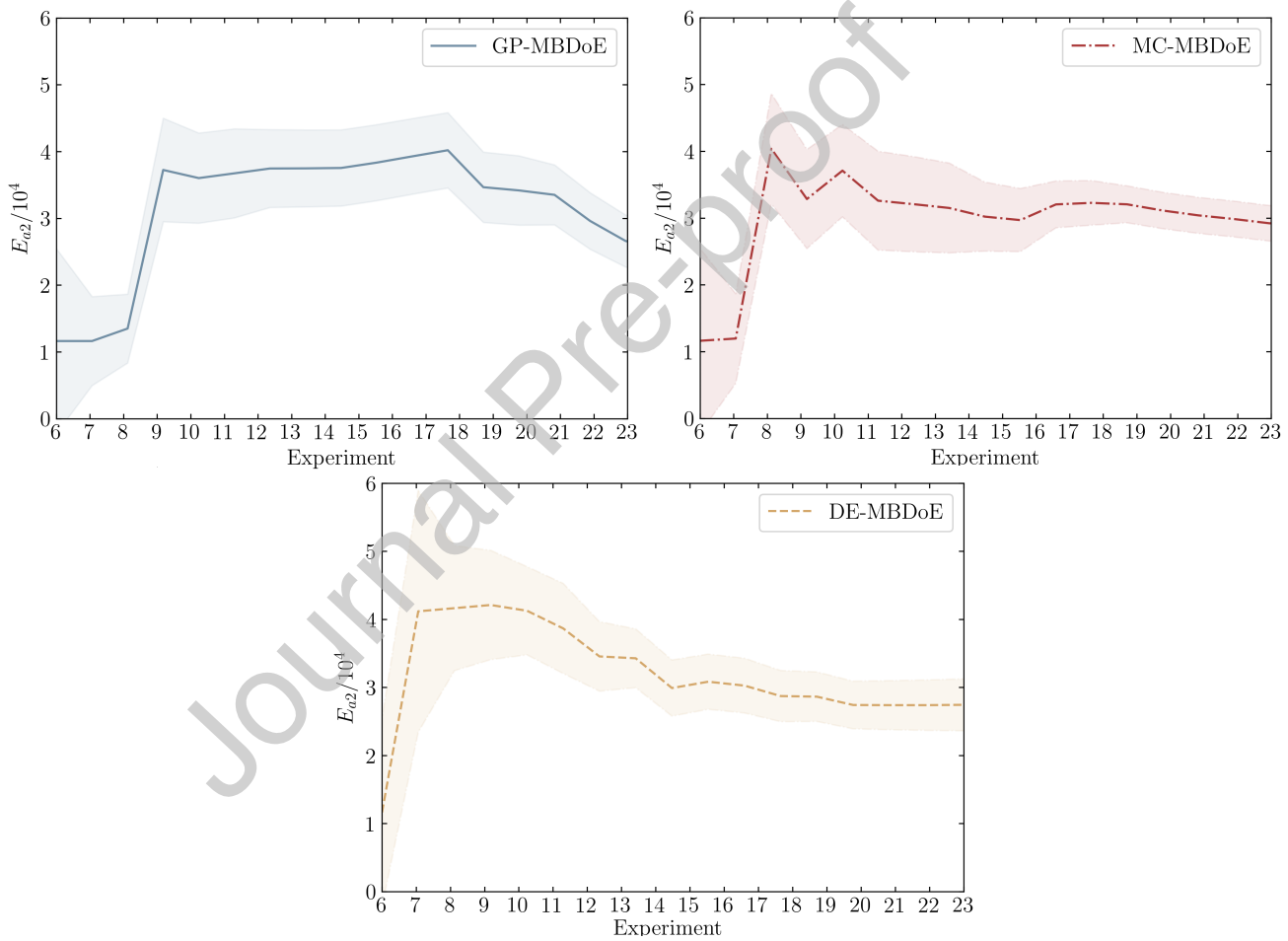
Model Parameters	Estimated Values	95% t (final)	$t_{ref}$ (final)	95% t (initial)	$t_{ref}$ (initial)
$k_1^0$ ( $M^{-1}min^{-1}$ )	5.23	16.41	1.65	9.07	1.74
$E_{a1}$ (kJ/mol)	27.4	12.17	1.65	6.23	1.74
$k_2^0$ ( $M^{-1}min^{-1}$ )	0.33	2.92	1.65	1.87	1.74
$E_{a2}$ (kJ/mol)	33.5	2.73	1.65	0.51*	1.74
$k_3^0$ ( $M^{-1}min^{-1}$ )	0.54	18.44	1.65	4.43	1.74
$E_{a3}$ (kJ/mol)	32.8	15.91	1.65	6.67	1.74
$k_4^0$ ( $M^{-1}min^{-1}$ )	1.41	2.22	1.65	1.28*	1.74
$E_{a4}$ (kJ/mol)	43.6	2.62	1.65	$9 \times 10^{-9}$ *	1.74

Table 11: Case study 2, MC-MBDoE. Parameter estimation results in terms of estimated values, 95% confidence intervals and t-values.

Model Parameters	Estimated Values	95% t (final)	$t_{ref}$ (final)	95% t (initial)	$t_{ref}$ (initial)
$k_1^0$ ( $M^{-1}min^{-1}$ )	4.45	28.56	1.65	9.07	1.74
$E_{a1}$ (kJ/mol)	19.70	17.44	1.65	6.23	1.74
$k_2^0$ ( $M^{-1}min^{-1}$ )	0.28	3.95	1.65	1.87	1.74
$E_{a2}$ (kJ/mol)	27.6	3.35	1.65	0.51*	1.74
$k_3^0$ ( $M^{-1}min^{-1}$ )	0.47	16.54	1.65	4.43	1.74
$E_{a3}$ (kJ/mol)	33.0	15.97	1.65	6.67	1.74
$k_4^0$ ( $M^{-1}min^{-1}$ )	1.32	1.8	1.65	1.28*	1.74
$E_{a4}$ (kJ/mol)	43.8	1.94	1.65	$9 \times 10^{-9}$ *	1.74

Table 12: Case study 2, DE-MBDoE. Parameter estimation results in terms of estimated values, 95% confidence intervals and t-values.

Model Parameters	Estimated Values	95% t (final)	$t_{ref}$ (final)	95% t (initial)	$t_{ref}$ (initial)
$k_1^0$ ( $M^{-1}min^{-1}$ )	4.31	23.71	1.65	9.07	1.74
$E_{a1}$ (kJ/mol)	24.2	17.16	1.65	6.23	1.74
$k_2^0$ ( $M^{-1}min^{-1}$ )	0.33	2.45	1.65	1.87	1.74
$E_{a2}$ (kJ/mol)	27.6	2.11	1.65	0.51*	1.74
$k_3^0$ ( $M^{-1}min^{-1}$ )	0.42	4.22	1.65	4.43	1.74
$E_{a3}$ (kJ/mol)	39.9	5.31	1.65	6.67	1.74
$k_4^0$ ( $M^{-1}min^{-1}$ )	1.98	1.327*	1.65	1.28*	1.74
$E_{a4}$ (kJ/mol)	34.3	1.44*	1.65	$9 \times 10^{-9}$ *	1.74

Figure 11: Case Study 2: estimation profiles of  $E_{a2}$  for the different designs. The shaded uncertainty area is represented by the square root of the 95% confidence intervals.

As in the previous case, the parameters are updated after each new sample is acquired from experiments. The parameter  $E_{a2}$  is illustrated in Fig. 11 and initially does not pass the  $t$ -test. The behaviour of the parameters for each design appears to be similar, with the MC-MBDoE and DE-MBDoE having a smoother behaviour of the estimates as the parametric uncertainty reduces. Nevertheless, the parameter estimates are not asymptotically stable for MC-MBDoE and GP-MBDoE. This behaviour could be potentially be caused by

the structural mismatch that is present in the models.

The result here is similar with Case study 1, where the proposed GP-MBDoE performs better than the MC-MBDoE, but with feasible designs.

Table 13: Case study 2, 95%  $\chi^2$ ,  $\chi_{ref}^2 = 132.14$ .

Method	$\chi^2$
GP-MBDoE	1805
MC-MBDoE	3892
DE-MBDoE	3079

In this case study the Table 13 shows that the GP-MBDoE managed to outperform the standard robust MBDoE methods (MC-MBDoE and DE-MBDoE), performing better in terms of their predicting capabilities. Additionally, the GP-MBDoE managed to remain feasible in all the experiments.

For completeness purposes, we have added the design variables for all the methods (GP-MBDoE, MC-MBDoE, DE-MBDoE) in Appendix C.

## 5. Discussion

In this section, an additional discussion regarding the results of the case studies is added. In the case studies, the proposed method (GP-MBDoE) was compared with two alternative methods, a method that deals with parametric only uncertainty (MC-MBDoE) and a disturbance based estimation method (DE-MBDoE). The GP-MBDoE has the main advantage: it can deal with potential model-mismatch and have as good performance as the standard MBDoE techniques in terms of information metrics. The non-parametric nature of the GPs offers the opportunity to approximate the physical system with a small number of data-points locally with the help of trust regions. This nice property comes with additional experiments' expense due to the conservative nature of the trust regions. This method's main computational burden is the training of the second Gaussian process for the objective function. An arbitrary large amount of data-points can be used for this surrogate, as only in-silico data points are needed. This can be an issue as the number of the design variables and model parameters increases; however, approximated methods have been proposed, including sparse GPs and generally variational methods for the training [62]. In this work, the computational times needed are on average 42 and 121 CPU-sec with 10 multi-starts for each GP and 10 multi-starts for the optimization for case study 1 and case study 2, respectively. The computational times for the GP-MBDoE, MC-MBDoE and DE-MBDoE are given in Table 14. Please note that a preliminary set of experiments is required, particularly we need  $n_u + 1$  data points to compute the initial Gaussian process as the covariance of the input data points needs to be invertible.

The MC-MBDoE method, on the other hand, can effectively deal with parametric uncertainty and has proven to be quite effective [60, 35]. Nevertheless, even in the absence of model mismatch, such an approach may have several issues: Despite the simplicity of such approaches, neither robust optimization nor stochastic programming addresses the ambiguity of the uncertainty set, i.e. the issue is to have a reliable representation of the parametric uncertainty region and the uncertainty for the confidence region itself. This issue could be solved by applying 1) an initial conservative design to account for small number of data and then 2) the standard MBDoE techniques in a receding horizon fashion [20] and updating the uncertainty set. In this paper,

this method's computational time is high due to the Monte Carlo approach that is followed, specifically 250 CPU-sec and 500 CPU-sec. Nevertheless, as mentioned, more computationally efficient methods have been proposed in the literature to solve optimal control problems with parametric uncertainty.

The DE-MBDoE, on the other hand, can be seen as a subcategory of the GP-MBDoE (as a constant value disturbance model is used) without the trust regions. The DE-MBDoE effectively can approximate the model mismatch with a constant value similarly to the well-known disturbance rejection techniques [38]. This method can be very effective mainly when the DE-MBDoE is applied in a receding horizon fashion as the constant value is updated very often. The disadvantages here are 1) the constant value may not represent the model mismatch and 2) the method may be over-confident, where regions far from the current point may be explored. This method is the most computationally efficient, as it does not require any MC, resulting in less than 20 CPU-sec for the first case study and less than 62 CPU-sec for the second case study.

Table 14: Computational times for each case study and method

Method	Case Study 1:	Case Study 2:
	CPU time(sec)	CPU time(sec)
GP-MBDoE	42	121
MC-MBDoE	250	500
DE-MBDoE	20	62

## 6. Conclusions

In this article a new methodology for the safe MBDoE based on GP and trust regions in the presence of structural uncertainty. GP-MBDoE was proposed and discussed; The optimal design of an experiment for improving parameter estimation is a particular form of optimization problem that can be very effective, but both optimality and feasibility of the designed experiment are important issues to consider. The proposed strategy allows to restrict the feasible space maintaining the optimal design in the feasible (safe) region of the state variables. Two simulated (in-silico) case studies have been proposed to assess the effectiveness of the new technique. In the first case study, the methodology was applied to a model of a chemical reaction in a flow reactor, where the assumed mechanism is inadequate to represent the system. The constraint tightening strategy allows estimating the parametric set describing the approximate model as well as run the experiments in the safe feasible region. In the second case study, dealing with a larger problem of chemical parallel reactions (SnAr), where the model assumptions are correct, however there is a presence of a disturbance that is dependent on the design variables due to equipment malfunction. The proposed methodology (GP-MBDoE) managed to offer predictive capabilities similar (Case study 1) or better (Case study 2) compared to the standard robust MBDoE methods (MC-MBDoE & DE-MBDoE), whilst resulting in feasible experimental designs. Future work will aim at further improving the proposed approach by using methods related to Bayesian calibration to calibrate the approximated model itself [28] via a Bayesian framework. The parameter of the kinetic model can be seen as calibration parameters and estimated through their posterior distribution, which provides a non-asymptotic approach for uncertainty quantification. Additionally, the structural mismatch can be potentially be diagnosed [48] and different structure could be potential be found using artificial neural networks [49].

## Acknowledgment

This project has received funding from the EPSRC grant projects (EP/R032807/1) is gratefully acknowledged.

## References

- [1] E. Arcari, L. Hewing, M. Schlichting, M. N. Zeilinger, 2019. Dual stochastic MPC for systems with parametric and structural uncertainty. arXiv xxx, 1–13.
- [2] E. Arcari, L. Hewing, M. Schlichting, M. N. Zeilinger, 2019. Dual stochastic mpc for systems with parametric and structural uncertainty. arXiv preprint arXiv:1912.10114.
- [3] H. Arellano-Garcia, T. Barz, B. Dorneanu, V. S. Vassiliadis, 2020. Real-time feasibility of nonlinear model predictive control for semi-batch reactors subject to uncertainty and disturbances. *Computers and Chemical Engineering* 133, 106529.
- [4] S. P. Asprey, S. Macchietto, 2002. Designing robust optimal dynamic experiments. *Journal of Process Control* 12 (4), 545–556.
- [5] F. Berkenkamp, M. Turchetta, A. P. Schoellig, A. Krause, 2017. Safe model-based reinforcement learning with stability guarantees. *Advances in Neural Information Processing Systems 2017-Decem (Nips)*, 909–919.
- [6] D. Bonvin, C. Georgakis, C. C. Pantelides, M. Barolo, M. A. Grover, D. Rodrigues, R. Schneider, D. Dochain, 2016. Linking Models and Experiments. *Industrial and Engineering Chemistry Research* 55 (25), 6891–6903.
- [7] E. Bradford, L. Imsland, 2018. Economic Stochastic Model Predictive Control Using the Unscented Kalman Filter. *IFAC-PapersOnLine* 51 (18), 417–422.
- [8] E. Bradford, L. Imsland, 2018. Stochastic Nonlinear Model Predictive Control Using Gaussian Processes. *2018 European Control Conference, ECC 2018*, 1027–1034.
- [9] E. Bradford, L. Imsland, D. Zhang, E. A. del Rio Chanona, 2020. Stochastic data-driven model predictive control using gaussian processes. *Computers & Chemical Engineering* 139, 106844.
- [10] D. G. Brown, J. Boström, 2016. Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? *Journal of Medicinal Chemistry* 59 (10), 4443–4458.
- [11] G. Buzzi-Ferraris, F. Manenti, 2009. Kinetic models analysis. *Chemical Engineering Science* 64 (5), 1061 – 1074.
- [12] C. Cartis, N. I. M. Gould, P. L. Toint, 2011. Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. *Mathematical Programming* 127 (2), 245–295.
- [13] A. R. Conn, K. Scheinberg, L. N. Vicente, 2005. Introduction to derivative-free optimization.
- [14] M. P. Deisenroth, D. Fox, C. E. Rasmussen, 2015. Gaussian Processes for Data-Efficient Learning in Robotics and Control. *IEEE Transactions on pattern analysis and machine intelligence* 37 (2).
- [15] E. A. del Rio Chanona, P. Petsagkourakis, E. Bradford, J. E. A. Graciano, B. Chachuat, 2021. Real-time optimization meets bayesian optimization and derivative-free optimization: A tale of modifier adaptation. *Computers & Chemical Engineering* 147, 107249.



- [16] M. Farina, L. Giulioni, R. Scattolini, 2016. Stochastic linear Model Predictive Control with chance constraints - A review. *Journal of Process Control* 44, 53–67.  
URL <http://dx.doi.org/10.1016/j.jprocont.2016.03.005>
- [17] G. Franceschini, S. Macchietto, 2008. Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science* 63 (19), 4846–4872.
- [18] P. I. Frazier, 2018. A tutorial on bayesian optimization.
- [19] F. Galvanin, M. Barolo, F. Bezzo, S. Macchietto, 2010. A backoff strategy for model-based experiment design under parametric uncertainty. *AIChE Journal* 56 (8), 2088–2102.
- [20] F. Galvanin, M. Barolo, G. Pannocchia, F. Bezzo, 2012. Online model-based redesign of experiments with erratic models: A disturbance estimation approach. *Computers & Chemical Engineering* 42, 138–151.
- [21] A. Geletu, M. Klöppel, H. Zhang, P. Li, 2013. Advances and applications of chance-constrained approaches to systems optimisation under uncertainty. *International Journal of Systems Science* 44 (7), 1209–1232.
- [22] A. Girard, C. Rasmussen, J. Q. Candela, R. Murray-Smith, 2003. Gaussian process priors with uncertain inputs application to multiple-step ahead time series forecasting. In: S. Becker, S. Thrun, K. Obermayer (Eds.), *Advances in Neural Information Processing Systems*. Vol. 15. MIT Press, pp. 545–552.
- [23] W. K. Hastings, 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57 (1), 97–109.
- [24] L. Hewing, J. Kabzan, M. N. Zeilinger, 2020. Cautious model predictive control using gaussian process regression. *IEEE Transactions on Control Systems Technology* 28 (6), 2736–2743.
- [25] L. Hewing, K. P. Wabersich, M. Menner, M. N. Zeilinger, 2020. Learning-Based Model Predictive Control: Toward Safe Learning in Control. *Annual Review of Control, Robotics, and Autonomous Systems* 3 (1), 269–296.
- [26] C. A. Hone, N. Holmes, G. R. Akien, R. A. Bourne, F. L. Muller, 2017. Rapid multistep kinetic model generation from transient flow data. *Reaction Chemistry and Engineering* 2 (2), 103–108.
- [27] D. R. Jones, M. Schonlau, W. J. Welch, 1998. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization* 13 (4), 455–492.
- [28] M. C. Kennedy, A. O'Hagan, 2001. Bayesian calibration of computer models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 63 (3), 425–464.
- [29] J. Ko, D. J. Klein, D. Fox, D. Haehnel, 2007. GP-UKF: Unscented Kalman Filters with Gaussian Process prediction and observation models. *IEEE International Conference on Intelligent Robots and Systems*, 1901–1907.
- [30] S. Körkel, E. Kostina, H. Bock, J. Schlöder, 2004. Numerical methods for optimal control problems in design of robust optimal experiments for nonlinear dynamic processes. *Optimization Methods and Software* 19 (3-4), 327–338.
- [31] W. Meeker, L. Escobar, 1995. Teaching about approximate confidence regions based on maximum likelihood estimation. *American Statistician* 49 (1), 48–53.
- [32] A. Mesbah, 2016. Stochastic model predictive control: An overview and perspectives for future research.
- [33] A. Mesbah, S. Streif, 2015. Probabilistic approach to robust optimal experiment design with chance constraints. *IFAC-PapersOnLine* 28 (8), 100–105.
- [34] A. Mesbah, S. Streif, R. Findeisen, R. D. Braatz, 2014. Stochastic nonlinear model predictive control with probabilistic constraints. *Proceedings of the American Control Conference*, 2413–2419.

- [35] P. Nimmegeers, S. Bhonsale, D. Telen, J. V. Impe, 2020. Optimal experiment design under parametric uncertainty: a comparison of a sensitivities based approach versus a polynomial chaos based stochastic approach. *Chemical Engineering Science* 221, 115651.
- [36] S. Olofsson, M. P. Deisenroth, R. Misener, 2018. Design of experiments for model discrimination hybridising analytical and data-driven approaches. 35th International Conference on Machine Learning, ICML 2018 9, 6259–6269.
- [37] E. Pan, P. Petsagkourakis, M. Mowbray, D. Zhang, A. del Rio-Chanona, 2020. Constrained model-free reinforcement learning for process optimization.
- [38] G. Pannocchia, J. B. Rawlings, 2003. Disturbance models for offset-free model-predictive control. *AIChE Journal* 49 (2), 426–437.
- [39] J. A. Paulson, T. L. M. Santos, A. Mesbah, 2019. Mixed stochastic-deterministic tube MPC for offset-free tracking in the presence of plant-model mismatch. *Journal of Process Control* 83, 102–120.
- [40] N. D. Perić, R. Paulen, M. E. Villanueva, B. Chachuat, 2018. Set-membership nonlinear regression approach to parameter estimation. *Journal of Process Control* 70, 80–95.
- [41] P. Petsagkourakis, W. P. Heath, J. Carrasco, C. Theodoropoulos, 2021. Robust stability of barrier-based model predictive control. *IEEE Transactions on Automatic Control* 66 (4), 1879–1886.
- [42] P. Petsagkourakis, W. P. Heath, C. Theodoropoulos, 2020. Stability analysis of piecewise affine systems with multi-model predictive control. *Automatica* 111, 108539.
- [43] P. Petsagkourakis, I. Sandoval, E. Bradford, D. Zhang, E. del Rio-Chanona, 2020. Reinforcement learning for batch bioprocess optimization. *Computers & Chemical Engineering* 133, 106649.
- [44] P. Petsagkourakis, I. O. Sandoval, E. Bradford, F. Galvanin, D. Zhang, E. A. del Rio-Chanona, 2020. Chance constrained policy optimization for process control and optimization. arXiv preprint arXiv:2008.00030.
- [45] L. Pronzato, E. Walter, 1985. Robust experiment design via stochastic approximation. *Mathematical Biosciences* 75 (1), 103 – 120.
- [46] M. Quaglio, E. S. Fraga, E. Cao, A. Gavriilidis, F. Galvanin, 2018. A model-based data mining approach for determining the domain of validity of approximated models. *Chemometrics and Intelligent Laboratory Systems* 172, 58–67.
- [47] M. Quaglio, E. S. Fraga, F. Galvanin, 2018. Constrained model-based design of experiments for the identification of approximated models. *IFAC-PapersOnLine* 51 (15), 515 – 520, 18th IFAC Symposium on System Identification SYSID 2018.
- [48] M. Quaglio, E. S. Fraga, F. Galvanin, 2020. A diagnostic procedure for improving the structure of approximated kinetic models. *Computers & Chemical Engineering* 133, 106659.
- [49] M. Quaglio, L. Roberts, M. S. Bin Jaapar, E. S. Fraga, V. Dua, F. Galvanin, 2020. An artificial neural network approach to recognise kinetic models from experimental data. *Computers & Chemical Engineering* 135, 106759.
- [50] M. Quaglio, C. Waldron, A. Pankajakshan, E. Cao, A. Gavriilidis, E. S. Fraga, F. Galvanin, 2019. An online reparametrisation approach for robust parameter estimation in automated model identification platforms. *Computers & Chemical Engineering*, 1–15.
- [51] C. Rasmussen, C. Williams, Jan. 2006. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, USA.

- [52] A. Reddy, G. Mukkula, R. Paulen, 2017. Model-based design of optimal experiments for nonlinear systems in the context of guaranteed parameter estimation. *Computers & Chemical Engineering* 99, 198–213.
- [53] C. R. Rojas, J. S. Welsh, G. C. Goodwin, A. Feuer, 2007. Robust optimal experiment design for system identification. *Automatica* 43 (6), 993 – 1008.
- [54] W. C. Rooney, L. T. Biegler, 2001. Design for model parameter uncertainty using nonlinear confidence regions. *AIChE Journal* 47 (8), 1794–1804.
- [55] B. Scholkopf, A. J. Smola, 2001. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA.
- [56] G. A. F. Seber, 2015. *Nonlinear Regression Models*. Springer International Publishing, Cham, pp. 117–128.
- [57] N. Srinivas, A. Krause, S. Kakade, M. Seeger, 2010. Gaussian process optimization in the bandit setting: No regret and experimental design. *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, 1015–1022.
- [58] S. Streif, F. Petzke, A. Mesbah, R. Findeisen, R. D. Braatz, 2014. Optimal experimental design for probabilistic model discrimination using polynomial chaos. Vol. 19. *IFAC*.
- [59] Y. Sui, A. Gotovos, J. W. Burdick, A. Krause, 2015. Safe exploration for optimization with Gaussian processes. *32nd International Conference on Machine Learning, ICML 2015* 2, 997–1005.
- [60] D. Telen, P. Nimmegeers, J. V. Impe, 2018. Uncertainty in optimal experiment design: comparing an online versus offline approaches. *IFAC-PapersOnLine* 51 (2), 771–776.
- [61] D. Telen, D. Vercammen, F. Logist, J. Van Impe, 2014. Robustifying optimal experiment design for nonlinear, dynamic (bio)chemical systems. *Computers & Chemical Engineering* 71, 415–425.
- [62] M. K. Titsias, 2009. Variational learning of inducing variables in sparse Gaussian processes. In: D. van Dyk, M. Welling (Eds.), *Journal of Machine Learning Research*. Vol. 5 of *Proceedings of Machine Learning Research*. PMLR, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, pp. 567–574.  
URL <http://proceedings.mlr.press/v5/titsias09a.html>
- [63] J. Umlauft, T. Beckers, S. Hirche, 2018. Scenario-based Optimal Control for Gaussian Process State Space Models. *2018 European Control Conference, ECC 2018*, 1386–1392.
- [64] R. Van Der Merwe, E. A. Wan, 2001. The square-root unscented Kalman filter for state and parameter-estimation. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* 6, 3461–3464.
- [65] M. Wang, R. S. Risuleo, E. W. Jacobsen, V. Chotteau, H. Hjalmarsson, 2020. Identification of nonlinear kinetics of macroscopic bio-reactions using multilinear Gaussian processes. *Computers & Chemical Engineering* 133, 106671.
- [66] J. S. Welsh, H. Kong, 2011. Robust experiment design through randomisation with chance constraints. *IFAC Proceedings Volumes* 44 (1), 13197 – 13202, 18th IFAC World Congress.
- [67] N. Wiener, 1938. The homogeneous chaos. *American Journal of Mathematics* 60 (4), 897–936.

## Appendix A. Uncertainty propagation with Gaussian processes

GP has found a success in the literature due to its attribute to estimate not only the expected value of a function (mean) but also the variance. However, the expressions for the posterior distribution of  $f_i$  in (13) hold only for a deterministic input  $\mathbf{x}^*$ . In the case that  $\mathbf{x}^*$  is random variable, then the expression (13) do not hold and for the case that  $\mathbf{x}^* \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}^*}, \boldsymbol{\Sigma}_{\mathbf{x}^*})$  the posterior is :

$$p(f_i(\mathbf{x}^*)|\boldsymbol{\mu}_{\mathbf{x}^*}, \boldsymbol{\Sigma}_{\mathbf{x}^*}, \mathcal{D}_i) = \int p(f_i(\mathbf{x}^*)|\mathbf{x}^*, \mathcal{D}_i)p(\mathbf{x}^*)d\mathbf{x}^*, \quad (\text{A.1})$$

which is computationally intractable.

### Appendix A.1. Gaussian Approximation: Formulation

To accommodate this issue, an analytical Gaussian approximation has been found, where only the mean ( $\hat{m}_i(\boldsymbol{\mu}_{\mathbf{x}^*}, \boldsymbol{\Sigma}_{\mathbf{x}^*})$ ) and variance ( $\hat{\Sigma}_i(\boldsymbol{\mu}_{\mathbf{x}^*}, \boldsymbol{\Sigma}_{\mathbf{x}^*})$ ) of  $f_i(\mathbf{x}^*)|\boldsymbol{\mu}_{\mathbf{x}^*}, \boldsymbol{\Sigma}_{\mathbf{x}^*}, \mathcal{D}_i$  is computed. These are obtained employing the law of iterated expectations and law of conditional variances [22]

$$\begin{aligned} \hat{m}_i(\boldsymbol{\mu}_{\mathbf{x}^*}, \boldsymbol{\Sigma}_{\mathbf{x}^*}) &= \mathbb{E}_{\mathbf{x}^*} \left( \mathbb{E}_{f_i^*} (f_i^*(\mathbf{x}^*)|\mathbf{x}^*) \right) = \mathbb{E}_{\mathbf{x}^*} \left( m_i(\mathbf{x}^*) \right) \\ \hat{\Sigma}_i(\boldsymbol{\mu}_{\mathbf{x}^*}, \boldsymbol{\Sigma}_{\mathbf{x}^*}) &= \mathbb{E}_{\mathbf{x}^*} \left( \mathbb{V}_{f_i^*} (f_i^*(\mathbf{x}^*)|\mathbf{x}^*) \right) + \mathbb{V}_{\mathbf{x}^*} \left( \mathbb{E}_{f_i^*} (f_i^*(\mathbf{x}^*)|\mathbf{x}^*) \right) \\ &= \mathbb{E}_{\mathbf{x}^*} \left( \Sigma_i(\mathbf{x}^*) \right) + \mathbb{V}_{\mathbf{x}^*} \left( m_i(\mathbf{x}^*) \right). \end{aligned} \quad (\text{A.2})$$

The expressions in (A.2) can analytically be found for the case of squared exponential. In the case of general case of kernels, Taylor expansion has been proposed to find a close form expression of (A.2). First order Taylor expansion is applied around  $\boldsymbol{\mu}_{\mathbf{x}^*}$  for  $m_i(\cdot, \cdot)$ :

$$\begin{aligned} \mathbb{E}_{\mathbf{x}^*} \left( m_i(\mathbf{x}^*) \right) &\approx \mathbb{E}_{\mathbf{x}^*} \left( m_i(\boldsymbol{\mu}_{\mathbf{x}^*}) + \frac{\partial m_i(\mathbf{x}^*)}{\partial \mathbf{x}^*} \Bigg|_{\boldsymbol{\mu}_{\mathbf{x}^*}}^T (\mathbf{x}^* - \boldsymbol{\mu}_{\mathbf{x}^*}) \right) = m_i(\boldsymbol{\mu}_{\mathbf{x}^*}) \\ \mathbb{V}_{\mathbf{x}^*} \left( m_i(\mathbf{x}^*) \right) &\approx \mathbb{V}_{\mathbf{x}^*} \left( m_i(\boldsymbol{\mu}_{\mathbf{x}^*}) + \frac{\partial m_i(\mathbf{x}^*)}{\partial \mathbf{x}^*} \Bigg|_{\boldsymbol{\mu}_{\mathbf{x}^*}}^T (\mathbf{x}^* - \boldsymbol{\mu}_{\mathbf{x}^*}) \right) = \frac{\partial m_i(\mathbf{x}^*)}{\partial \mathbf{x}^*} \Bigg|_{\boldsymbol{\mu}_{\mathbf{x}^*}}^T \boldsymbol{\Sigma}_{\mathbf{x}^*} \frac{\partial m_i(\mathbf{x}^*)}{\partial \mathbf{x}^*} \Bigg|_{\boldsymbol{\mu}_{\mathbf{x}^*}} \end{aligned} \quad (\text{A.3})$$

and second order Taylor for the variance  $\Sigma_i(\cdot, \cdot)$ :

$$\begin{aligned} \Sigma_i(\mathbf{x}^*) &\approx \Sigma_i(\boldsymbol{\mu}_{\mathbf{x}^*}) + \frac{\partial \Sigma_i(\mathbf{x}^*)}{\partial \mathbf{x}^*} \Bigg|_{\boldsymbol{\mu}_{\mathbf{x}^*}}^T (\mathbf{x}^* - \boldsymbol{\mu}_{\mathbf{x}^*}) + \frac{1}{2} (\mathbf{x}^* - \boldsymbol{\mu}_{\mathbf{x}^*})^T \frac{\partial^2 \Sigma_i(\mathbf{x}^*)}{\partial \mathbf{x}^* \partial \mathbf{x}^{*T}} \Bigg|_{\boldsymbol{\mu}_{\mathbf{x}^*}} (\mathbf{x}^* - \boldsymbol{\mu}_{\mathbf{x}^*}) \\ \mathbb{E}_{\mathbf{x}^*} \left( \Sigma_i(\mathbf{x}^*) \right) &\approx \Sigma_i(\boldsymbol{\mu}_{\mathbf{x}^*}) + \frac{1}{2} \text{trace} \left\{ \frac{\partial^2 \Sigma_i(\mathbf{x}^*)}{\partial \mathbf{x}^* \partial \mathbf{x}^{*T}} \Bigg|_{\boldsymbol{\mu}_{\mathbf{x}^*}} \boldsymbol{\Sigma}_{\mathbf{x}^*} \right\}. \end{aligned} \quad (\text{A.4})$$

Now, the approximated mean  $\hat{m}_i$  and variance  $\hat{\Sigma}_i$  can be computed by substituting (A.3)&(A.4) to (A.2):

$$\begin{aligned} \hat{m}_i(\boldsymbol{\mu}_{\mathbf{x}^*}, \boldsymbol{\Sigma}_{\mathbf{x}^*}) &= m_i(\boldsymbol{\mu}_{\mathbf{x}^*}) \\ \hat{\Sigma}_i(\boldsymbol{\mu}_{\mathbf{x}^*}, \boldsymbol{\Sigma}_{\mathbf{x}^*}) &= \Sigma_i(\boldsymbol{\mu}_{\mathbf{x}^*}) + \frac{1}{2} \text{trace} \left\{ \frac{\partial^2 \Sigma_i(\mathbf{x}^*)}{\partial \mathbf{x}^* \partial \mathbf{x}^{*T}} \Bigg|_{\boldsymbol{\mu}_{\mathbf{x}^*}} \boldsymbol{\Sigma}_{\mathbf{x}^*} \right\} + \frac{\partial m_i(\mathbf{x}^*)}{\partial \mathbf{x}^*} \Bigg|_{\boldsymbol{\mu}_{\mathbf{x}^*}}^T \boldsymbol{\Sigma}_{\mathbf{x}^*} \frac{\partial m_i(\mathbf{x}^*)}{\partial \mathbf{x}^*} \Bigg|_{\boldsymbol{\mu}_{\mathbf{x}^*}}. \end{aligned} \quad (\text{A.5})$$

The approximated solution of the Gaussian approximation can be advantageous for the the propagation of uncertainty. In this paper approximation (A.5) will be used to propagate the uncertainty of the model parameters to the Fisher information matrix.

Appendix A.2. Gaussian Approximation: Simple Numerical Example

Let a simple function  $f(x) = \sin(x)$  be assumed as the ‘true’ model, with  $x \in \mathbb{R}$  and the collected data are corrupted by a Gaussian noise  $\epsilon_f$  with standard deviation 0.01,  $y = f(x) + \epsilon_f$ . A GP is used to approximate the data points, specifically a squared-exponential (SE) covariance function is used and its hyperparameters are fitted using maximum likelihood estimation (with multi-starts to avoid local optima). The mean  $m$  and variance  $\Sigma$  of the GP for each  $x$  are illustrated in Fig A.12.

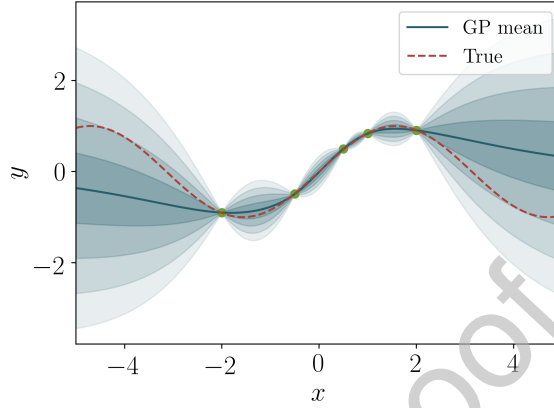


Figure A.12: The ground (red dashed line) is the function that the GP approximates given the available data. The GP predictions for the mean ( $\mu$ ) is the blue line with variance  $\Sigma$ , and the shaded areas represents the area between  $m \pm \sqrt{\Sigma}$ ,  $m \pm \sqrt{\Sigma}$ ,  $m \pm 3\sqrt{\Sigma}$ ,  $m \pm 4\sqrt{\Sigma}$  and  $m \pm 10\sqrt{\Sigma}$ .

The shaded areas show the  $m + r\sqrt{\Sigma}$  for  $r = 1, 2, 3, 4$  and 10. If the test point  $x^*$  is deterministic then the prediction for  $f(x^*)$  would follow a Gaussian distribution with mean  $m$  and variance  $\Sigma$ ; however such computation is not correct if  $x^*$  is a random variable. Let’s assume that  $x^* \sim \mathcal{N}(0.1, 0.01)$  then Fig. A.13 depicts 3 different approaches of computing the posterior density function; 1) Monte-Carlo approximation: This distribution is approximated using Monte-Carlo sampling (1000 samples), 2) Mean value approximation: The computation of the posterior ignores the stochastic nature of the test point  $x^*$  and computes the posterior  $m$  and  $\Sigma$  using (13), and 3) Gaussian approximation: The posterior distribution is approximated using the analytical Gaussian approximation (A.5).

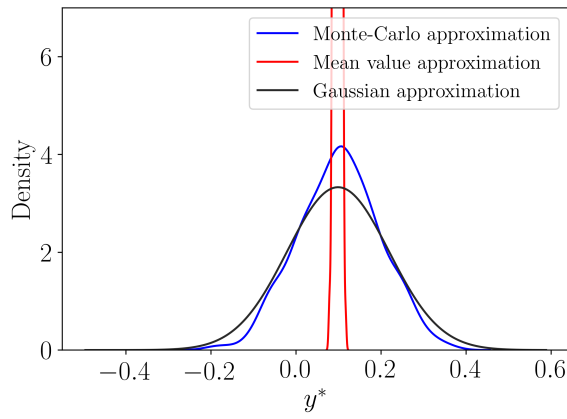


Figure A.13: Comparison of prediction methods for random input  $x^*$ . The posterior distribution is evaluated with the different approximation techniques; Monte-Carlo approximation (blue), Mean value approximation (red) and Gaussian approximation (black).

It is apparent that the posterior distribution is not Gaussian and that the approximation as a Gaussian distribution leads to error, even if the first mean and variance are exactly the same. The Mean value Approximation can be very conservative or over-confident (like in this example) as it neglects into account the stochastic nature of the input.

## Appendix B. Nomenclature

Table B.15: Nomenclature

Symbol	Description
$\mathbf{x}$	States
$\mathbf{u}$	Design variables
$\mathbf{w}$	External disturbances
$\nu$	Random noise
$\hat{\boldsymbol{\theta}}$	Model parameters
$\boldsymbol{\theta}$	Model parameters
$\hat{(\cdot)}$	Approximate quantity ( $\cdot$ )
$n_x$	Dimension of $\mathbf{x}$
$\mathbf{V}_{\boldsymbol{\theta}}$	Variance-covariance of the parameters $\boldsymbol{\theta}$
$\mathcal{D}$	Data available
$\mathcal{L}(\boldsymbol{\theta} \mathcal{D})$	The joint probability of the prediction-observation mismatch in all data points ( $\mathcal{D}$ ) for the parameter $\boldsymbol{\theta}$
$\mathbb{E}$	Expectation
$\mathbb{P}$	Probability
$\mathbb{V}$	Variance
$\mathbb{U}$	Constraints for the design variables
$\dot{\mathbf{x}}$	derivative with respect to time of variable $\mathbf{x}$
$g_j$	Constraint $j$
$\mathcal{GP}$	Gaussian Process (Distribution over functions)

Table B.16: Nomenclature (continued)

Symbol	Description
$\mu_i$	Prior mean for GP
$k_i$	Kernel function for GP of $i^{th}$ state
$K$ or $K(\mathbf{X}, \mathbf{X})$	Covariance (or Gram) matrix
$m_i$	Posterior mean for GP of $i^{th}$ state
$\Sigma_i$	Variance of $i^{th}$ state
$\mathcal{N}(\cdot, \cdot)$	Normal distribution defined by two quantities, i.e. mean and variance-covariance
$\sim$	Distributed according to; example: $x \sim \mathcal{N}(\mu, \sigma^2)$
$\ \cdot\ _{k_i}$	Norm in the associated Reproducing Kernel Hilbert Space
$\mathcal{H}_k$	Reproducing kernel Hilbert space (RKHS)
$\mathcal{R}$	Radius of trust region
$\rho$	Percentage error of model/Measurement
$t_1, t_2$	Adaptation parameters of trust region
$\eta_1, \eta_2$	Threshold for <i>rho</i>
$t$	$t$ value for $t$ -test
$\chi^2$	Chi square value
$\bar{\alpha}$	Probability of constraints violation
$1 - \alpha$	Confidence level
$\alpha_J$	Exploration term for MBDoE
$k_j^r$	Kinetic reaction parameter of $j^{th}$ reaction
$k_j^0$	Pre-exponential of $j^{th}$ reaction
$E_{a_j}$	Activation energy of $j^{th}$ reaction
$F$	Flowrate
$c$	Concentrations
$T$	Temperature

### Appendix C. Design variables for Case studies

In this section the design variables for the case study 1 and 2 are given.



## Appendix C.1. Case study 1: Design Variables

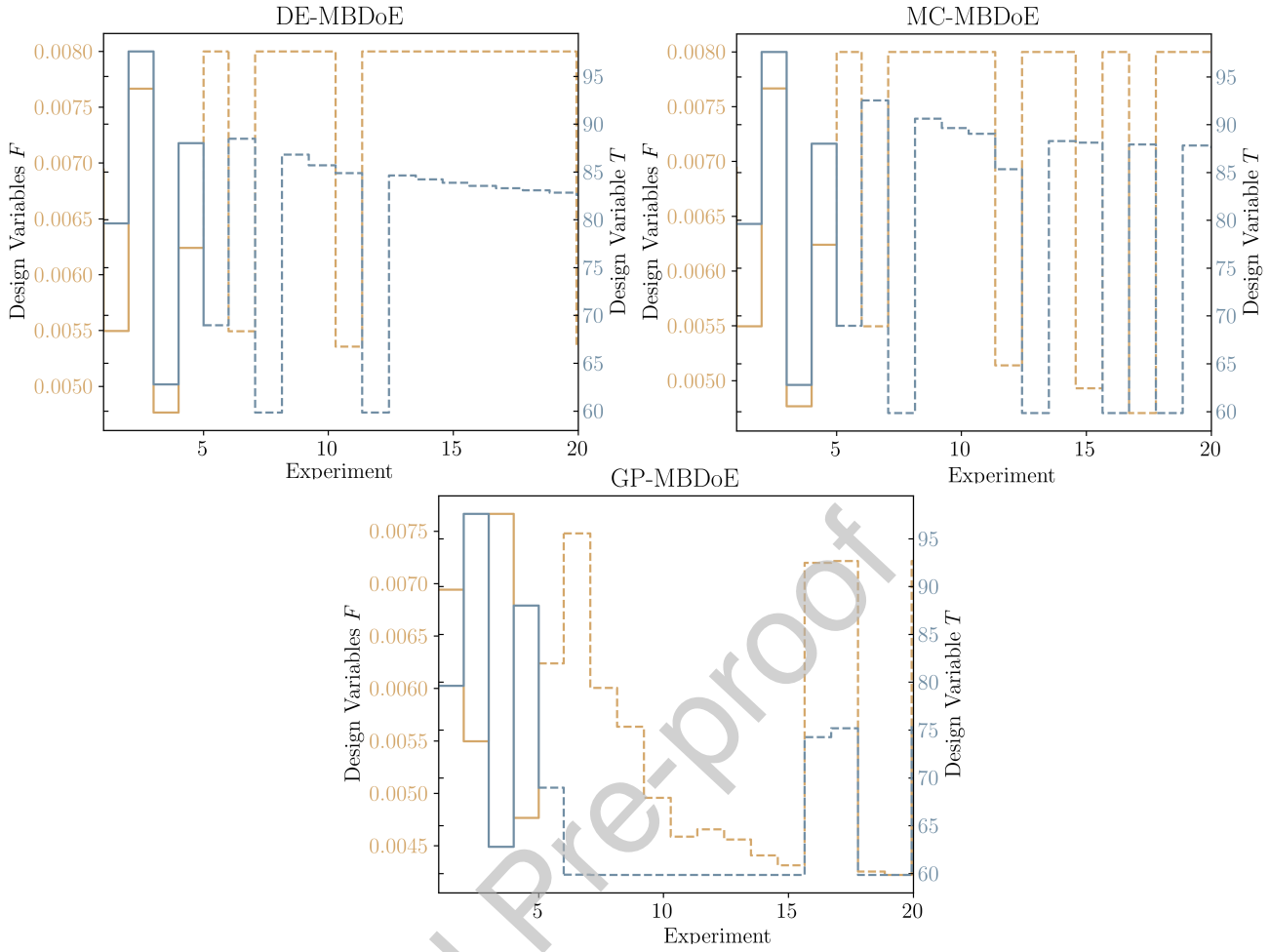


Figure C.14: Case Study 2: Computed design variables for the different designs.

It is clear from Fig. C.14, that the 3 frameworks result in 3 fundamentally different strategies. The DE-MBDoE converges to high temperatures and flowrates, and the design variables from MC-MBDoE oscillate. Conversely, the GP-MBDoE appears to focus on smaller values as the number of experiments increases.

## Appendix C.2. Case study 2: Design Variables

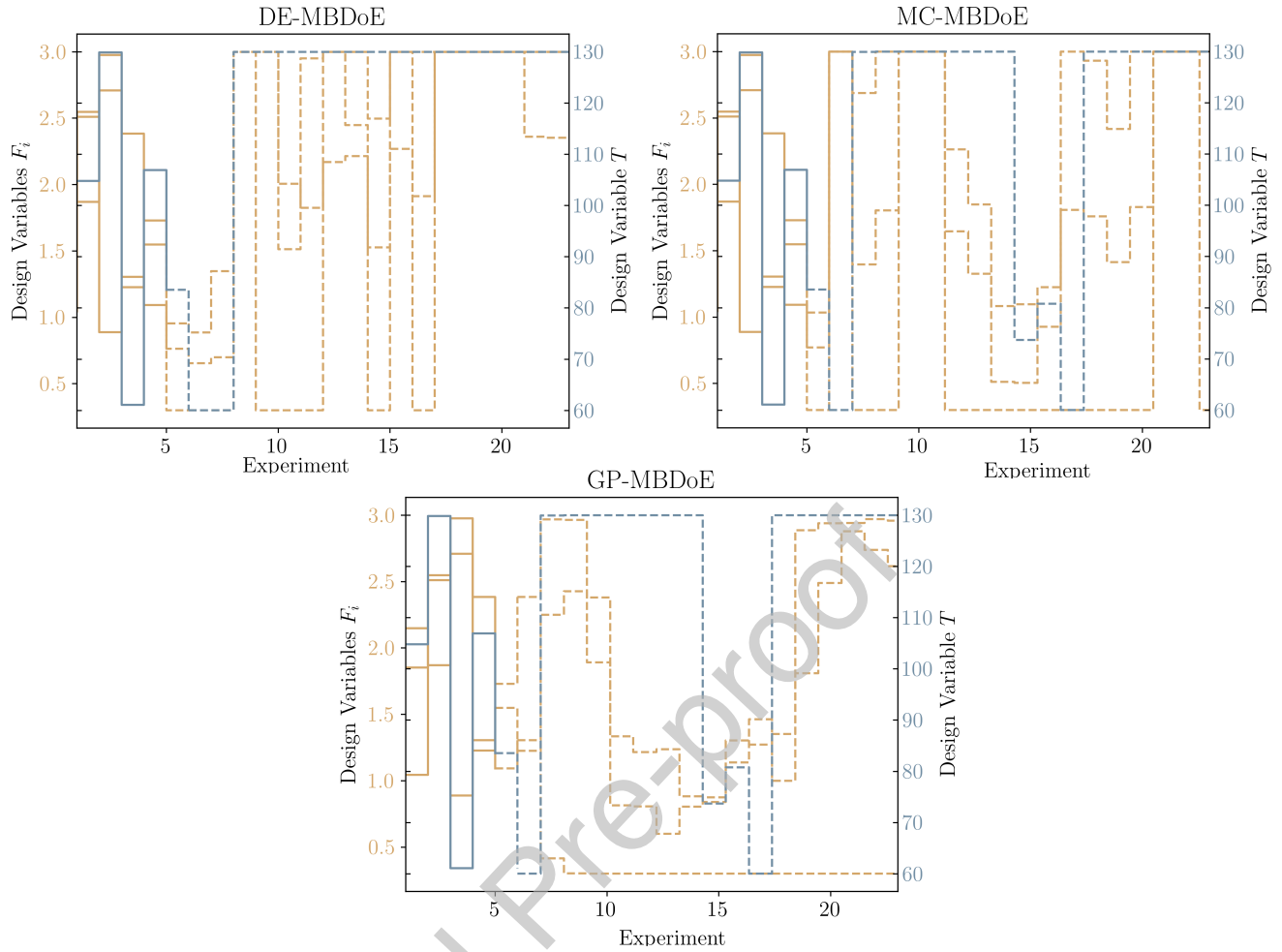


Figure C.15: Case Study 2: Computed design variables for the different designs.

The results for the design variables of the second case study are depicted in C.15. All the methods seem to prefer the use of high temperature as the number of experiments increases. Specifically, both MC-MBDoE and GP-MBDoE provide the same designs for the temperature. On the other hand, the flowrates are designed differently, where MC-MBDoE results in oscillations and the flowrates of DE-MBDoE converge to the highest value. GP-MBDoE selects to operate in the lowest flowrate for  $F_1$  and high flowrates for  $F_2$  and  $F_3$ .

**Declaration of interests**

The authors declare no conflict of interest.

Journal Pre-proof

**CRedit authorship contribution statement:**

**P. Petsagkourakis:**

Conceptualization, Software, Methodology, Formal analysis, Writing - original draft, Visualization

**F. Galvanin:**

Conceptualization, Methodology, Writing - review & editing, Supervision

Journal Pre-proof