



**HAL**  
open science

## Sequential design of multi-fidelity computer experiments: maximizing the rate of stepwise uncertainty reduction

Rémi Stroh, Julien Bect, Séverine Demeyer, Nicolas Fischer, Damien Marquis,  
Emmanuel Vazquez

### ► To cite this version:

Rémi Stroh, Julien Bect, Séverine Demeyer, Nicolas Fischer, Damien Marquis, et al.. Sequential design of multi-fidelity computer experiments: maximizing the rate of stepwise uncertainty reduction. *Technometrics*, 2022, 64 (2), pp.199-209. 10.1080/00401706.2021.1935324 . hal-02902333v3

**HAL Id: hal-02902333**

**<https://centralesupelec.hal.science/hal-02902333v3>**

Submitted on 15 Jul 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0  
International License

# Sequential design of multi-fidelity computer experiments: maximizing the rate of stepwise uncertainty reduction

Rémi STROH<sup>\*†</sup>      Julien BECT<sup>\*</sup>      Séverine DEMEYER<sup>†</sup>  
Nicolas FISCHER<sup>†</sup>      Damien MARQUIS<sup>‡</sup>      Emmanuel VAZQUEZ<sup>\*</sup>

<sup>\*</sup> Université Paris-Saclay, CNRS, CentraleSupélec,  
Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France  
Email: [firstname.lastname@l2s.centralesupelec.fr](mailto:firstname.lastname@l2s.centralesupelec.fr)

<sup>†</sup> Département sciences des données et incertitudes

<sup>‡</sup> Département comportement au feu et sécurité incendie

Laboratoire national de métrologie et d'essais, Trappes, France  
Email: [firstname.lastname@lne.fr](mailto:firstname.lastname@lne.fr)

## Abstract

This article deals with the sequential design of experiments for (deterministic or stochastic) multi-fidelity numerical simulators, that is, simulators that offer control over the accuracy of simulation of the physical phenomenon or system under study. Accurate simulations usually entail a high computational effort, while coarse simulations are obtained at a lower cost. The cost can be measured, e.g., by the run time of the simulator or the financial cost of the computing resources. In this setting, simulation results obtained at several levels of fidelity can be combined in order to estimate quantities of interest (the optimal value of the output, the probability that the output exceeds a given threshold. . .) in an efficient manner. We propose a new Bayesian sequential strategy called Maximal Rate of Stepwise Uncertainty Reduction (MR-SUR), that selects additional simulations to be performed by maximizing the ratio between the expected reduction of uncertainty and the cost of simulation. This generic strategy unifies several existing methods, and provides a principled approach to develop new ones. We assess its performance on several examples, including a computationally intensive problem of fire safety analysis where the quantity of interest is the probability of exceeding a tenability threshold during a building fire.

*Keywords:* Multi-fidelity, Computer experiments, Sequential design of experiments, Gaussian process emulator, Meta-model, Surrogate model, Stochastic simulator

# Contents

1	Introduction . . . . .	3
2	Gaussian process models for multi-fidelity . . . . .	5
2.1	The auto-regressive model for deterministic simulators . . . . .	5
2.2	The additive model for deterministic simulators . . . . .	6
2.3	Extension to stochastic simulators . . . . .	7
3	Sequential design of experiment for multi-fidelity . . . . .	8
3.1	Existing methods . . . . .	8
3.2	Stepwise uncertainty reduction . . . . .	9
3.3	Maximum rate of stepwise uncertainty reduction . . . . .	10
4	Numerical results . . . . .	12
4.1	Setup of the experiments . . . . .	13
4.2	A one-dimensional example . . . . .	14
4.3	Random damped harmonic oscillator . . . . .	15
4.4	A fire safety example . . . . .	18
5	Conclusion . . . . .	20
	 SUPPLEMENTARY MATERIAL . . . . .	 28
SM-1	Introduction . . . . .	28
SM-2	Nonsequential designs in multi-fidelity . . . . .	28
SM-3	A new SUR criterion . . . . .	29
SM-3.1	Criterion definition and result statement . . . . .	29
SM-3.2	A useful identity . . . . .	30
SM-3.3	Proof of Proposition 1 . . . . .	31
SM-4	Additional information regarding the examples . . . . .	32
SM-4.1	One-dimensional example . . . . .	32
SM-4.2	Random damped harmonic oscillator . . . . .	32

# 1 Introduction

In the domain of computer experiments, *multi-fidelity* refers to the idea of combining results from numerical simulations at different levels of accuracy: high-fidelity simulations correspond to more accurate but, in general, more expensive computations. As a representative example of a multi-fidelity simulator, consider the case of a partial differential equation (PDE) solver based on a finite element or finite volume method: the accuracy of the numerical solution depends, among other things, on the fineness of the discretization. High-fidelity results are obtained when the mesh size is small.

Conceptually, a simulator is viewed in this article as a *black box* with inputs and outputs. The parameter that controls the level of accuracy/fidelity—the mesh size in the case of a PDE solver—is one of the inputs of this black box, alongside others such as design or control variables and environmental variables (see, e.g., Santner et al., 2018). Examples of multi-fidelity simulators can be found in virtually all areas of engineering and science, including aeronautics (Forrester et al., 2007), fire safety (Demeyer et al., 2017), electromechanics (Hage Hassan et al., 2014), electromagnetism (Koziel et al., 2013), hæmodynamics (Perdikaris and Karniadakis, 2016) and many more.

The cost of a simulation is often quantified by the run time of the simulator, or the financial cost of the computing resources, but other measures of cost can be more relevant in some contexts—for instance energy consumption or carbon emission in a green computing perspective (see, e.g., Kansal and Chana, 2012). We use henceforth the word “cost” as a generic term to refer to whichever measure of cost is relevant, in the application at hand, to express the available budget of simulations.

When the objective is to estimate a particular quantity of interest (QoI), such as the optimal value of the design variables (optimization problem) or the probability that the outputs belong to a prescribed “safe region” (reliability problem), multi-fidelity makes it possible to obtain a good approximation of the QoI with a computational effort lower than what would have been necessary if high-fidelity simulations only had been carried out. This cost reduction is achieved through the joint use of multi-fidelity models, which allow simulation results obtained at different levels of fidelity to be combined, and multi-fidelity designs of experiments (DoE). See Giselle Fernández-Godino et al. (2016) for a review that covers both aspects. This article addresses the problem of constructing, sequentially, a multi-fidelity DoE targeting a given QoI.

We adopt a Bayesian point of view following the line of research initiated by Sacks et al. (1989), where prior belief about the simulator is modeled using a Gaussian process (see Santner et al., 2018, and references therein). The Bayesian approach provides a rich framework for the construction of sequential DoE, which has been abundantly relied upon

in previous works dealing with the case of single-fidelity simulators, where the cost of a simulation is assumed to be independent of the value of the input variables (see, e.g., Kushner, 1964; Mockus et al., 1978; Jones et al., 1998; Ranjan et al., 2008; Villemonteix et al., 2009; Picheny et al., 2010; Bect et al., 2012; Chevalier et al., 2014). In this framework, sequential designs are usually constructed by means of a sampling criterion—also called infill criterion, acquisition function or merit function. The value of the criterion indicates whether a particular point in the input space is promising or not. The expected improvement (EI) criterion (Jones et al., 1998) is a popular example of such a sampling criterion. The extension of the Bayesian approach to sequential DoE in a multi-fidelity setting is based on two ingredients: 1) the construction of prior models for simulators with adjustable fidelity; 2) the construction of sampling criteria that take the variable cost of simulations into account.

For the case of deterministic multi-fidelity simulators, Gaussian process-based models have already been proposed in the literature (Kennedy and O’Hagan, 2000; Qian et al., 2006; Qian and Wu, 2008; Le Gratiet, 2013; Le Gratiet and Garnier, 2014; Picheny and Ginsbourger, 2013; Tuo et al., 2014). Extensions to combinations of computer experiments and physical experiments (Qian and Wu, 2008) and to stochastic simulators (Stroh et al., 2017a) have been considered as well.

Sampling criteria for single-fidelity sequential designs do not reflect a crucial feature of multi-fidelity simulators: the cost of a run depends on the value of the inputs (in particular on the one that controls the fidelity of simulation). Various methods that take into account a variable cost of simulation have been proposed for particular cases: single-objective unconstrained optimization (Huang et al., 2006; Swersky et al., 2013; He et al., 2017) and global approximation (Xiong et al., 2013; Le Gratiet and Cannamela, 2015) notably, and more recently reliability analysis and contour estimation (Marques et al., 2018, 2020; Yi et al., 2021).

In this article, we provide a general principled methodology to construct sequential DoE for multi-fidelity simulators and, more generally, for simulators where the cost of a simulation depends on the value of the inputs. The methodology is applicable to any QoI, and builds on the *Stepwise Uncertainty Reduction* (SUR) principle (see, e.g., Villemonteix et al., 2009; Bect et al., 2012; Chevalier et al., 2014; Bect et al., 2019, and references therein), which unifies many of the aforementioned sequential DoE for the fixed-cost case. More precisely, for the variable-cost case, we propose the *Maximal Rate of Stepwise Uncertainty Reduction* (MR-SUR) principle, which consists in constructing a sequential design by maximizing, at each step, the ratio between the expected reduction of uncertainty (to be defined more precisely later on) and the simulation cost.

The article is organized as follows. Section 2 reviews Gaussian process modeling for

deterministic simulators, and discusses some possible extensions to (normally distributed) stochastic simulators. Section 3 first reviews both existing methods of sequential design for multi-fidelity simulators and the SUR principle for fixed-cost simulators, and then presents the MR-SUR principle and its relations with some existing sequential DoE. Finally, Section 4 illustrates the method and assesses its performance through several academic examples, including a computationally intensive problem of fire safety analysis where the quantity of interest is the probability of exceeding a tenability threshold during a building fire.

## 2 Gaussian process models for multi-fidelity

We consider a computer simulator with input variables  $\mathbf{u} \in \mathbb{U} \subset \mathbb{R}^d$  and one or several scalar outputs, which are generally obtained after some post-processing steps (e.g., an aerodynamic drag in a CFD model). Moreover, we consider that the accuracy, or *fidelity*, of the computer simulation can be tuned using a parameter  $\delta$  that ranges in a discrete or continuous set  $\mathbb{T}$ . For instance,  $\delta$  can be the mesh size in a finite element or finite volume method. Such a parameter will be called *fidelity parameter* and can be viewed as an additional input of the simulator. We denote by  $\mathbf{x} = (\mathbf{u}, \delta) \in \mathbb{X}$  the aggregated vector of inputs, with  $\mathbb{X} = \mathbb{U} \times \mathbb{T}$ , and we assume from now on that the output is scalar.

### 2.1 The auto-regressive model for deterministic simulators

The so-called *auto-regressive model* of Kennedy and O'Hagan (2000) assumes a deterministic simulator with a finite number  $S$  of levels of increasing fidelity. Let  $\delta_1, \dots, \delta_S$  denote the corresponding values of the fidelity parameter and set  $\mathbb{T} = \{\delta_1, \dots, \delta_S\}$ . The simulator is then modeled by a Gaussian process  $\xi$  on  $\mathbb{X} = \mathbb{U} \times \mathbb{T}$ , defined through an auto-regressive relationship between successive levels:

$$\begin{cases} \xi(\mathbf{u}, \delta_1) = \eta_1(\mathbf{u}), \\ \xi(\mathbf{u}, \delta_s) = \rho_{s-1} \xi(\mathbf{u}, \delta_{s-1}) + \eta_s(\mathbf{u}), \quad 1 < s \leq S, \end{cases} \quad (1)$$

where  $\eta_1, \dots, \eta_S$  are  $S$  mutually independent Gaussian processes, and  $(\rho_s)_{1 \leq s < S} \in \mathbb{R}^{S-1}$ .

The model has been used in numerous applications, where the actual number  $S$  of levels is most often two (see, e.g., Qian et al., 2006; Forrester et al., 2007; Qian and Wu, 2008; Kuya et al., 2011; Brooks et al., 2011; Goh et al., 2013; Le Gratiet and Garnier, 2014; Le Gratiet and Cannamela, 2015; Elsayed, 2015; Demeyer et al., 2017), occasionally three (Perdikaris and Karniadakis, 2016, Section 3.2). In practice, the Gaussian processes  $\eta_s$  are chosen among a family of Gaussian processes indexed by parameters such as correlation

lengths, regularity parameters, etc., which are estimated from data (simulation results), by maximum likelihood for instance (see, e.g., Stein (1999)). Since the processes  $\eta_s$  are assumed independent, there must be enough simulation results at each level of fidelity, even at the possibly very expensive highest fidelity levels, to obtain good estimates of the parameters—which explains perhaps why this model is typically used with a small number of levels.

*Remark 1.* Some authors introduce more flexibility in the model by letting the coefficients  $\rho_s$  in Equation (1) depend on  $\mathbf{u}$  as well, for instance linearly (Qian et al., 2006; Le Gratiet and Garnier, 2014) or using another Gaussian process in a hierarchical manner (Qian and Wu, 2008).

## 2.2 The additive model for deterministic simulators

Another approach to building Gaussian process models for deterministic multi-fidelity simulators, which readily applies to the case where  $\mathbb{T}$  contains a continuum of levels of fidelity or a large number of ordered discrete levels of fidelity, has been proposed by Picheny and Ginsbourger (2013) and Tuo et al. (2014).

Assuming for simplicity that  $\mathbb{T} = [0, \infty)$ , with  $\delta = 0$  corresponding as in Tuo et al. (2014) to the highest—often unreachable—level of fidelity, a Gaussian process  $\xi$  over the product space  $\mathbb{X} = \mathbb{U} \times \mathbb{T}$  is defined in this approach as the sum of two independent parts:

$$\xi(\mathbf{u}, \delta) = \xi_0(\mathbf{u}) + \varepsilon(\mathbf{u}, \delta), \quad (2)$$

where  $\xi_0$  and  $\varepsilon$  are mutually independent Gaussian processes, and  $\varepsilon$  has zero mean and goes to zero in the mean-square sense when  $\delta \rightarrow 0$ . In other words,  $\text{var}(\varepsilon(\mathbf{u}, \delta)) \rightarrow 0$  when  $\delta \rightarrow 0$ , for all  $\mathbf{u}$ : as a consequence,  $\xi$  is a *non-stationary* Gaussian process on  $\mathbb{X} \subset \mathbb{R}^{d+1}$ . Under this decomposition,  $\xi_0$  represents the “ideal” version of the simulator, while  $\varepsilon$  represents numerical error.

In both articles,  $\xi_0$  is then assumed to be stationary, whereas the covariance function of  $\varepsilon$  is multiplicatively separable: for all  $\mathbf{u}, \mathbf{u}' \in \mathbb{U}$  and  $\delta, \delta' \in \mathbb{T}$ ,

$$\text{cov}(\varepsilon(\mathbf{u}, \delta), \varepsilon(\mathbf{u}', \delta')) = r(\delta, \delta') k(\mathbf{u}, \mathbf{u}'), \quad (3)$$

where  $k$  is a stationary covariance function on  $\mathbb{U}$ , and  $r$  is a non-stationary covariance function on  $\mathbb{T}$  such that  $r(\delta, \delta) \rightarrow 0$  when  $\delta \rightarrow 0$ . As an example of a suitable choice for  $r$ , consider the Brownian-type model proposed by Tuo et al. (2014):

$$r(\delta, \delta') = \min\{\delta, \delta'\}^L, \quad (4)$$

with  $L$  a positive parameter. Other choices are of course possible.

## 2.3 Extension to stochastic simulators

We turn now to the case of stochastic simulators, that is, simulators whose output is stochastic, as happens for instance when the computer program relies on a Monte Carlo method (see, e.g., Cochet et al., 2014). Extending the multi-fidelity Bayesian methodology of Sections 2.1 and 2.2 to stochastic simulators is not straightforward in general, since the output at a given input point  $\mathbf{x}_i = (\mathbf{u}_i, \delta_i) \in \mathbb{X}$  is now a random variable  $Z_i$ , the distribution of which is in general unknown and different at each point in  $\mathbb{X}$ . (Several runs at the same input point yield independent and identically distributed responses.)

We focus in this section on the simpler case where the output  $Z_i$  can be assumed to be normally distributed:

$$Z_i | \xi, \lambda \sim \mathcal{N}(\xi(\mathbf{x}_i), \lambda(\mathbf{x}_i)), \quad (5)$$

with mean  $\xi(\mathbf{x}_i)$  and variance  $\lambda(\mathbf{x}_i)$  possibly depending on the input point. In this setting, we propose to extend the multi-fidelity models of previous sections using independent prior distributions for  $\xi$  and  $\lambda$ , with either the autoregressive model of Section 2.1 or the additive model of Section 2.2 as a prior for  $\xi$ . Then, since the variance  $\lambda$  must have positive values, we suggest modeling  $\log(\lambda)$  by a Gaussian process  $\tilde{\lambda}$ , following Goldberg et al. (1998), Kersting et al. (2007), Boukouvalas and Cornford (2009) and others.

Under this type of model, the inference task—estimating the parameters of the Gaussian process models for  $\xi$  and  $\tilde{\lambda}$ , and then computing posterior distributions—becomes more difficult since neither  $\xi$  nor  $\tilde{\lambda}$  are directly observable. Goldberg et al. (1998) take a fully Bayesian approach using a time-consuming Monte-Carlo method. Other authors suggest optimization-based approaches, that simultaneously produce estimates the parameters of of both Gaussian processes and of the unobserved log-variances. In particular, Kersting et al. (2007) and Boukouvalas and Cornford (2009) propose a method called *most likely heteroscedastic GP*, stemming from the Expectation-Maximization (EM) algorithm (see also Marrel et al., 2012, for a similar algorithm), while Binois et al. (2018) use a more sophisticated joint maximization procedure with relaxation to obtain the joint MAP (maximum a posterior) estimator.

For the numerical experiments of this article (Sections 4.3 and 4.4) we will take a simpler route, assuming that the variance  $\lambda$  depends only on the fidelity level  $\delta$ —which is approximately true in the two examples we shall consider. In this setting, as long as the number of fidelity levels of interest is not too large, the value of the variance at these levels can be simply estimated jointly with the other parameters of the model; a general-purpose log-normal prior for the vector of variances is proposed by Stroh et al. (2017b).

## 3 Sequential design of experiment for multi-fidelity

### 3.1 Existing methods

In the literature of multi-fidelity, a variety of sequential design algorithms have been proposed. (See Supplementary Material for a review of *nonsequential* multi-fidelity designs, which can be used as initial designs for sequential ones.)

For instance, Forrester et al. (2007) suggest using the auto-regressive model of Kennedy and O'Hagan (2000) and a standard single-level sequential design at the highest level of fidelity to select input variables  $\mathbf{u} \in \mathbb{U}$  for the next experiment. Then, simulations at all levels of fidelity are run for the selected  $\mathbf{u}$ . Building on Forrester et al. (2007), Kuya et al. (2011) suggest a two-stage method: run a large number of simulations at the low-fidelity level, and then use a sequential design strategy to select simulations at the high-fidelity level. In a different spirit, Xiong et al. (2013) use Nested Latin Hypercube Sampling (NLHS) and suggest to double the number of simulations when going from level  $\delta^{(s)}$  to level  $\delta^{(s+1)}$ , until some cross-validation-based criterion is satisfied.

More interestingly in the context of this article, some methods have been proposed that explicitly take into account the simulation cost. This is typically achieved by crafting a sampling criterion that takes the form of a ratio between a term which measures the interest of a simulation at  $(\mathbf{u}, \delta)$ , and the cost of the simulation (Huang et al., 2006; Le Gratiet and Cannamela, 2015; He et al., 2017; Marques et al., 2018, 2020; Yi et al., 2021). For instance, He et al. (2017) propose a global optimization method using the Expected Quantile Improvement (EQI) of Picheny et al. (2013) and the multi-fidelity model of Tuo et al. (2014), and build a new sampling criterion corresponding to the ratio between the EQI sampling criterion and the simulation cost.

Outside the multi-fidelity literature, a similar idea has been proposed by Johnson (1960) to design sequential testing procedures, by Swersky et al. (2013) for multi-task optimization and by Poloczek et al. (2017) for multi-information source optimization. In the first two cases, the numerator of the criterion is the expected reduction of the entropy of the QoI, while in the third one it is equal to the Knowledge Gradient criterion (Frazier et al., 2009).

In this article, we propose a general methodology to build such sequential designs, which is not tied to a particular kind of model or QoI. The key idea is to measure the potential of a particular design point using the SUR framework, recalled in Section 3.2. The methodology itself, that we call MR-SUR, is presented in Section 3.3.

### 3.2 Stepwise uncertainty reduction

We recall here the principle of SUR strategies, introduced in the design of computer experiments by Vazquez and co-authors (see, e.g., Villemonteix et al., 2009; Bect et al., 2012). Given a Bayesian model of a simulator and an unknown QoI  $Q$ , that is, a particular feature of the simulator that we want to estimate, a SUR strategy is a Bayesian method for the construction of a sequence of evaluation locations  $\mathbf{X}_1, \mathbf{X}_2, \dots \in \mathbb{X}$  at which observations of the simulator will be taken in order to reduce the uncertainty on  $Q$ . (In this section,  $\mathbb{X}$  denotes a generic input space, not necessarily of the form  $\mathbb{X} = \mathbb{U} \times \mathbb{T}$ .)

The starting point of the construction of a SUR strategy is the definition of a statistic  $H_n$  measuring the residual uncertainty about  $Q$  given past observations  $Z_1, \dots, Z_n$ . Many choices for  $H_n$  are possible for any particular problem, but a natural requirement (Bect et al., 2019) is that  $H_n$  should be decreasing on average when  $n$  increases. For instance,  $H_n$  can be the posterior Shannon entropy of  $Q$  for a discrete QoI, or the posterior (total) variance for a scalar (or vector-valued) QoI. If  $Q$  is a function defined on  $\mathbb{X}$ , as will be the case in Section 4, a possible choice is

$$H_n = \mathbb{E}_n \left( \|Q - \hat{Q}_n\|_\mu^2 \right) = \int_{\mathbb{X}} \text{var}_n(Q(\mathbf{x})) \mu(d\mathbf{x}), \quad (6)$$

where  $\mu$  denotes a measure on  $\mathbb{X}$ ,  $\|h\|_\mu^2 = \int_{\mathbb{X}} h(\mathbf{x})^2 \mu(d\mathbf{x})$ ,  $\mathbb{E}_n$  (resp.  $\text{var}_n$ ) is the posterior expectation (resp. variance) given  $Z_1, \dots, Z_n$ , and  $\hat{Q}_n(\mathbf{x}) = \mathbb{E}_n(Q(\mathbf{x}))$ .

Then, given past observations,  $\mathbf{X}_{n+1}$  is chosen by minimizing the expectation of the future residual uncertainty:

$$\mathbf{X}_{n+1} = \underset{\mathbf{x} \in \mathbb{X}}{\text{argmin}} J_n(\mathbf{x}), \quad \text{with } J_n(\mathbf{x}) = \mathbb{E}_n(H_{n+1} \mid \mathbf{X}_{n+1} = \mathbf{x}), \quad (7)$$

where the expectation is with respect to the outcome  $Z_{n+1}$  of a new simulation at  $\mathbf{x} \in \mathbb{X}$ . Similarly, a batch of  $q$  evaluation points can be chosen by minimizing a multi-point criterion:

$$J_{n,q}(\mathbf{x}_1, \dots, \mathbf{x}_q) = \mathbb{E}_n(H_{n+q} \mid \mathbf{X}_{n+1} = \mathbf{x}_1, \dots, \mathbf{X}_{n+q} = \mathbf{x}_q) \quad (8)$$

(see, e.g., Chevalier et al., 2014; Azzimonti et al., 2021).

**Example.** Assume a stochastic multi-fidelity simulator defined over  $\mathbb{X} = \mathbb{U} \times \mathbb{T}$  as in Section 2.3, and consider the functional QoI defined on  $\mathbb{X}$  by

$$Q(\mathbf{x}) = \mathbb{P}\left(Z_{\mathbf{x}} > z^{\text{crit}} \mid \xi, \lambda\right) = \Phi\left(\frac{\xi(\mathbf{x}) - z^{\text{crit}}}{\sqrt{\lambda(\mathbf{x})}}\right), \quad (9)$$

where  $Z_{\mathbf{x}}$  denotes the outcome of a new simulation at  $\mathbf{x}$ ,  $z^{\text{crit}} \in \mathbb{R}$  is a given threshold, and  $\Phi$  the cdf of the standard normal distribution. Pick some reference level  $\delta^{\text{ref}} \in \mathbb{T}$  and consider the residual uncertainty

$$H_n = \int_{\mathbb{U}} \text{var}_n(Q(\mathbf{u}, \delta^{\text{ref}})) d\mathbf{u}, \quad (10)$$

which is a special case of Equation (6) with  $\mu$  equal to Lebesgue's measure on  $\mathbb{U}$  at fixed  $\delta = \delta^{\text{ref}}$ . Then, using computations similar to those of Chevalier et al. (2014), it can be proved that

$$J_n(\mathbf{x}) = \int_{\mathbb{U}} \left[ \Phi_2\left(a_n(\mathbf{x}'), a_n(\mathbf{x}'); \frac{k_n(\mathbf{x}', \mathbf{x}')}{v_n(\mathbf{x}')}\right) - \Phi_2\left(a_n(\mathbf{x}'), a_n(\mathbf{x}'); \frac{k_n(\mathbf{x}, \mathbf{x}')^2}{v_n(\mathbf{x})v_n(\mathbf{x}')}\right) \right] d\mathbf{u}', \quad (11)$$

where  $\mathbf{x}' = (\mathbf{u}', \delta^{\text{ref}})$ ,  $m_n$  (resp.  $k_n$ ) denotes the posterior mean (resp. covariance) of  $\xi$ ,  $v_n(\mathbf{x}) = \lambda(\mathbf{x}) + k_n(\mathbf{x}, \mathbf{x})$ ,  $a_n(\mathbf{x}) = (m_n(\mathbf{x}) - z^{\text{crit}}) / \sqrt{v_n(\mathbf{x})}$ , and  $\Phi_2(\cdot, \cdot; \rho)$  is the cdf of the standard bivariate normal distribution with correlation  $\rho$ . (For tractability, the variance function  $\lambda$  is assumed to be known in the computation of the criterion. In practice, the estimated variance function is plugged in the expression, and the integral over  $\mathbb{U}$  is approximated using a Monte Carlo method.)

*Remark 2.* See Supplementary Material for a proof of Equation (11), in a more general form which also allows for (possibly parallel) batches of evaluations as in (8), and integration with respect to a generic measure  $\mu$  as in Equation (6).

*Remark 3.* In the special case  $\lambda \equiv 0$  (deterministic simulator), corresponding to  $Q(\mathbf{x}) = \mathbb{1}_{\xi(\mathbf{x}) > z^{\text{crit}}}$ , the criterion (11) has been proposed by Bect et al. (2012) and computed by Chevalier et al. (2014). The general case, to the best of our knowledge, is new.

The reader is referred to, e.g., Villemonteix et al. (2009), Picheny et al. (2010), Chevalier et al. (2014), and Bect et al. (2019) for other examples of SUR criteria.

### 3.3 Maximum rate of stepwise uncertainty reduction

The proposed Maximum Rate of Stepwise Uncertainty Reduction (MR-SUR) strategy builds on the SUR strategy presented in Section 3.2. The goal is to achieve a balance between the (expected) reduction of uncertainty brought by new observations on the one hand, and the cost of these observations, usually measured by their computation time, on the other hand. Denote by  $C : \mathbb{X} \rightarrow \mathbb{R}_+$  the cost of an observation of the simulator, which in general depends mainly on the fidelity level  $\delta \in \mathbb{T}$ , but can also depend on the input  $\mathbf{u} \in \mathbb{U}$  in some applications. The MR-SUR strategy is then given by

$$\mathbf{X}_{n+1} = \operatorname{argmax}_{\mathbf{x} \in \mathbb{X}} \frac{H_n - J_n(\mathbf{x})}{C(\mathbf{x})} = \operatorname{argmax}_{\mathbf{x} \in \mathbb{X}} \frac{G_n(\mathbf{x})}{C(\mathbf{x})}, \quad (12)$$

where  $G_n(\mathbf{x}) = H_n - J_n(\mathbf{x})$  is the *expected uncertainty reduction* associated to a future observation at  $\mathbf{x} \in \mathbb{X}$ . This strategy boils down to a SUR strategy when  $C$  is constant.

A few special cases of MR-SUR strategies, adapted to particular models and estimation goals, have been proposed earlier in the literature. To the best of our knowledge, the oldest

example is the sequential testing method of Johnson (1960), where  $H_n$  is the posterior entropy of the location of faulty component in an electronic equipment—with a discrete distribution over all possible fault locations as the underlying model. More recently, Snoek et al. (2012), Swersky et al. (2013), and Poloczek et al. (2017) have proposed Bayesian optimization procedures of the MR-SUR type, for unconstrained global optimization problems with variable-cost evaluations, corresponding respectively, when the cost is constant, to the expected improvement (Mockus et al., 1978; Jones et al., 1998), IAGO (Villemon-teix et al., 2009) and KG (Frazier et al., 2009) algorithms—see Bect et al. (2019) for a presentation of EI and KG in the SUR framework. The first sequential design procedure of Le Gratiet and Cannamela (2015) can also be seen as an approximate MR-SUR strategy for the approximation problem, where  $H_n$  is the posterior integrated prediction variance. More recently, the CLoVER algorithm of Marques et al. (2018, 2020) provides an illustration of the MR-SUR principle for the problem of contour estimation.

*Remark 4.* Assuming that the cost depends only on the fidelity level  $\delta$ , the MR-SUR principle extends very naturally to batches of evaluations with a common cost  $C(\delta)$ , replacing the expected uncertainty  $J_n(\mathbf{x})$  in the numerator by  $J_{n,q}(\mathbf{x}_1, \dots, \mathbf{x}_q)$ , where  $q$  is the size of the batch (as in Equation (8)). The application of this idea to (synchronous) parallel computations—assuming that  $C(\delta)$  is the computation time—is left for future work.

To illustrate the MR-SUR principle, let us consider a simple simulated example, with  $\xi$  a Gaussian process on  $\mathbb{X} = \mathbb{U} \times \mathbb{T} = [-0.5, 0.5] \times [0, 1]$  such that  $\xi \mid m \sim \text{GP}(m, k)$ ,  $m \sim \mathcal{U}(\mathbb{R})$ , and  $k$  as in Section 2.2:

$$k((u, \delta), (u', \delta')) = \sigma_0^2 \mathcal{M}_{\nu_0} \left( \frac{|u - u'|}{\rho_0} \right) + \sigma_0^2 G \min \{ \delta, \delta' \}^L \mathcal{M}_{\nu_\varepsilon} \left( \frac{|u - u'|}{\rho_\varepsilon} \right),$$

where  $\mathcal{U}(\mathbb{R})$  denotes the (improper) uniform distribution on  $\mathbb{R}$  and  $\mathcal{M}_\nu$  the Matérn correlation function with regularity parameter  $\nu$ . The values  $m = 0$ ,  $\sigma_0 = 1$ ,  $G = 4$ ,  $L = 2$ ,  $\nu_0 = \nu_\varepsilon = 5/2$ ,  $\rho_0 = 0.3$ ,  $\rho_\varepsilon = 0.1$  are used in the simulations, and all the parameters except  $m$  are assumed to be known in this experiment. The cost function is defined as  $C(u, \delta) = 1/\delta$ . The QoI  $Q$  and the measure of residual uncertainty  $H_n$  are defined as in Equations (9)–(10), with  $z^{\text{crit}} = 0$ . Note that the level of highest fidelity  $\delta = 0$  is not observable in practice. A NLHS of size  $n = 12 + 6 + 6 + 3$  on the levels  $\delta = 1, 1/2, 1/5, 1/10$  is taken as the observed DoE, and the outputs  $Z_1, \dots, Z_n$  are simulated according to Equation (5) with constant variance  $\lambda = 0.4^2$ . We compute the functions  $J_n$ ,  $G_n$  and  $C$  over a  $100 \times 100$  regular grid on  $\mathbb{U} \times \mathbb{T}$ , to obtain Figures 1a and 1b.

Observe on Figure 1a that, for each cost value (corresponding to a fixed level of fidelity), there is a range of points that yield different values of the expected uncertainty reduction. Good observation points lie on the Pareto front (in solid black line), that is, the set of points

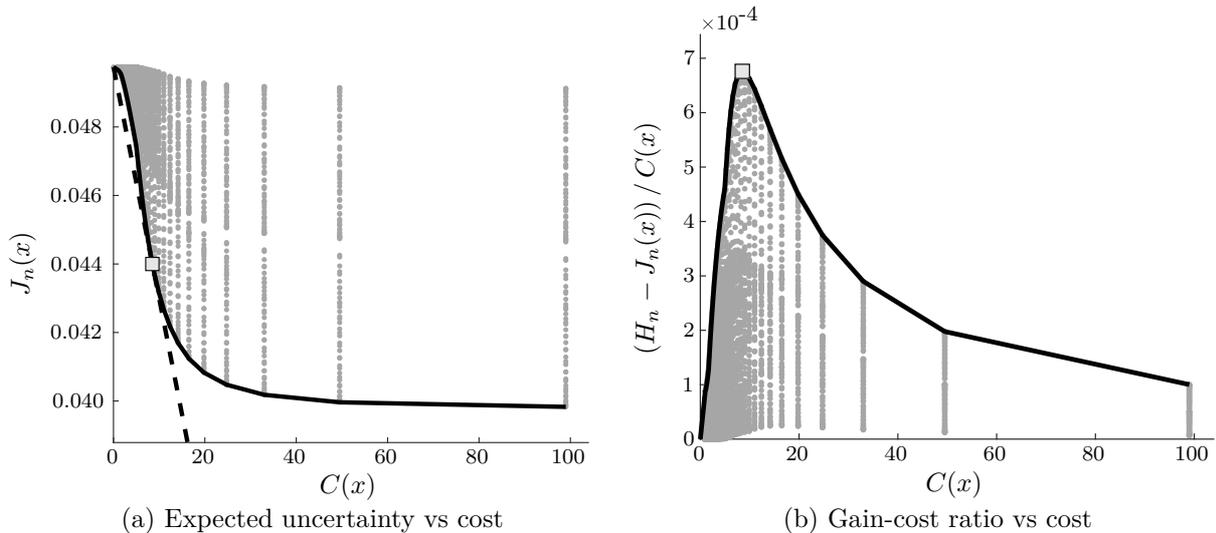


Figure 1: Illustration of the MR-SUR criterion. (a) Representation of possible designs in the  $(C, J)$  plane. (b) Representation in the  $(C, G/C)$  plane. Each gray point corresponds to one particular input point  $x = (u, \delta)$ , the solid line indicates Pareto-optimal points, and the square is the design selected by the MR-SUR criterion.

for which there is no larger expected uncertainty reduction at lower cost. The MR-SUR strategy selects an observation location that correspond to the maximum of the “slope” of this Pareto front.

Figure 2 shows the sequence of Pareto fronts as more observation points are added in the design using Equation (11). The horizontal axis now represents the total cost of the design, including the cost of the candidate point, so that the left-ends of the Pareto fronts are shifted (the top-left curve corresponds to the situation depicted on Figure 1a). This figure illustrates how the MR-SUR strategy progressively reduces the uncertainty about the QoI, by selecting at each step a pair  $(u, \delta)$  that makes a trade-off between cost and expected uncertainty reduction.

*Remark 5.* In the numerical experiments of Section 4, it will be assumed that the cost function  $C(x)$  is known beforehand. When it is not, it is usually estimated prior to the sequential design phase using a combination of expert knowledge and pilot runs (as described, e.g., by He et al., 2017). Alternatively, it could be estimated during the sequential design phase, using for instance (as suggested by Swersky et al., 2013) a second GP model for  $x \mapsto \log C(x)$ .

## 4 Numerical results

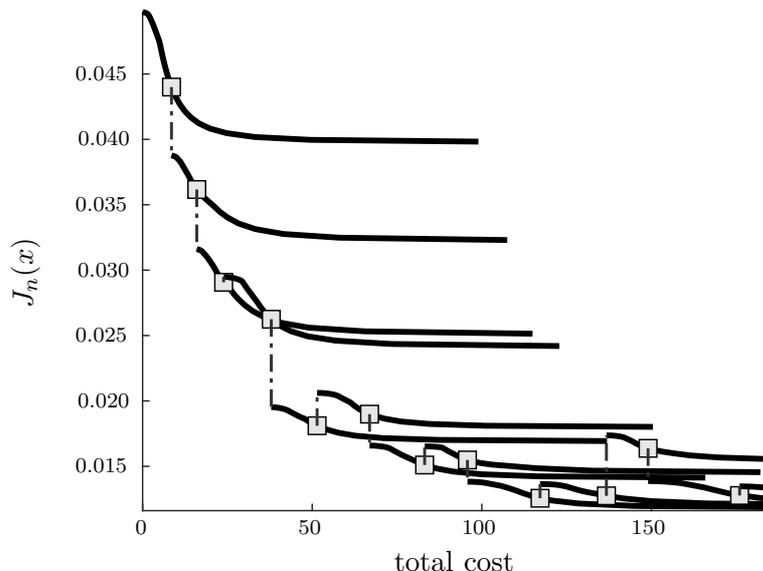


Figure 2: Sequence of Pareto fronts in  $(C, J)$ -space, as a function of the total cost of the design, in a particular run of the MR-SUR algorithm. On each curve, the square indicates the design selected by the MR-SUR criterion. A mixed black line connects the expected uncertainty at this point with the actual uncertainty obtained after the simulation has been run.

## 4.1 Setup of the experiments

In each example, we consider a multi-fidelity simulator for which the simulation cost  $C$  depends on  $\delta$  alone, and is assumed to be known. Some common features of all three numerical experiments are presented in this section.

*Initial DoE.* A nested Latin hypercube sample (NLHS) is used as an initial design. More specifically, we use the algorithm developed by Qian (2009), with an additional maximin optimization at each level to obtain better space-filling properties (see Stroh, 2018, Section 2.2.3 for details).

*GP modeling.* In each example, a multi-fidelity GP model of the type described in Section 2 is used. The posterior distribution of the parameters is initially sampled using an adaptive Metropolis-Hastings algorithm (Haario et al., 2001) and then updated at each iteration by sequential Monte Carlo (see, e.g., Chopin, 2002). More details about the particular GP model that is used, and the prior distribution on the parameters, are provided inside each example section.

*Optimization of the sampling criterion.* At each iteration of a SUR or MR-SUR strategy, a new simulation point is selected according to Equation (7) or Equation (12). This step involves an optimization of the SUR or MR-SUR criterion, which is carried out in the experiments of this article using a simple two-step approach: the criterion is first optimized by exhaustive search on a regular grid over  $\mathbb{U} \times \mathbb{T}$ , and then a local optimization is performed

starting from the best point in the grid. Other approaches have been proposed in the literature, that would be more efficient in higher-dimensional problems (see, e.g., Feliot et al., 2017).

*Other computational details.* All integrals are approximated by Monte-Carlo methods. SUR and MR-SUR criteria are evaluated using the Maximum A Posteriori (MAP) estimator of the parameters—obtained by local optimization from the best point in the MCMC/SMC sample—in a plug-in manner. (A fully Bayesian approach could be considered in principle, but would lead to much higher computational complexity.)

## 4.2 A one-dimensional example

Consider as a first (toy) example the two-level deterministic simulator defined for  $u \in [0; 1]$  and  $\delta \in \{1, 2\}$  by the analytical formulas (Forrester et al., 2007)

$$\begin{cases} f_1(u) = f(u, 1) = 0.5(6u - 2)^2 \sin(12u - 4) + 10(u - 0.5), \\ f_2(u) = f(u, 2) = (6u - 2)^2 \sin(12u - 4) + 10, \end{cases} \quad (13)$$

and assume that computing  $f_2$ —hereafter referred to as the “high fidelity” function—is four times more costly than computing  $f_1$ :  $C(2) = 1$  and  $C(1) = \frac{1}{4}$ . Note that the two functions are related by  $f_2(u) = 2f_1(u) - 20(u - 1)$ , which makes them perfect candidates for the autoregressive model presented in Section 2.1. The goal in this example is to estimate the set

$$\Gamma = \{f_2 > z^{\text{crit}}\} = \{u \in \mathbb{U}, f_2(u) \geq z^{\text{crit}}\}$$

with  $z^{\text{crit}} = 10$ . The performance of MR-SUR for this task will be compared with that of SUR strategies operating at the low-fidelity level only (LF-SUR) or at the high-fidelity level only (HF-SUR).

In this experiment, all three sequential strategies start with the same multi-fidelity initial design, and use the same Gaussian process prior and measure of uncertainty  $H_n$ . The initial design consists of six observations at the low-fidelity level and three at the high-fidelity level, for a total of  $n = 9$  observations, corresponding to an initial budget of  $6 \times \frac{1}{4} + 3 \times 1 = 4.5$  cost units. A supplementary budget of 9.0 cost units is assumed to be available for the sequential design. The autoregressive model of Section 2.1 is used, with Matérn 5/2 covariance functions and weakly informative priors on the parameters (see Supplementary Material for details). The uncertainty on  $\Gamma$  is quantified using the uncertainty measure (10). In the special case of a deterministic simulator, we have (cf. Remark 3)

$$H_n = \int_0^1 p_n(u) (1 - p_n(u)) du,$$

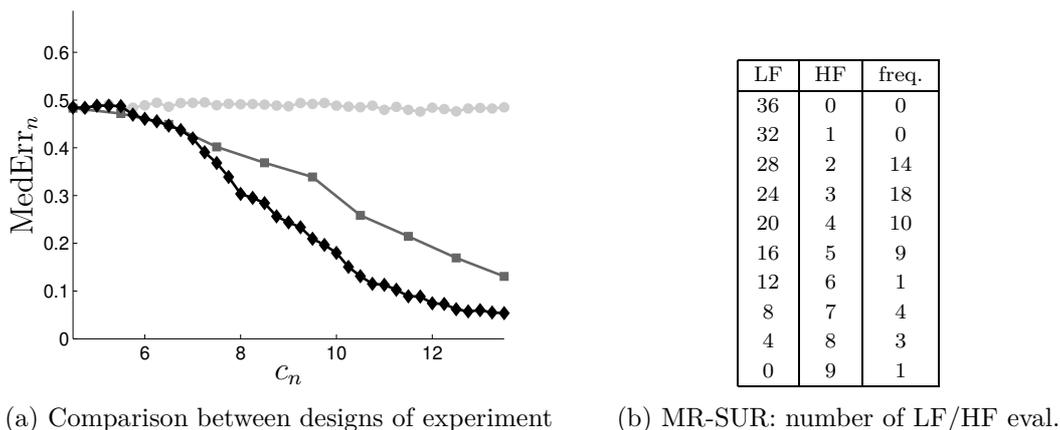


Figure 3: The one-dimensional experiment. (a) Median estimation error as a function of the cost. Light-gray disks: LF-SUR; dark-gray squares: HF-SUR; black diamonds: MR-SUR. (b) Number of LF/HF evaluations in the MR-SUR strategy. The last column indicates how many times, in the 60 repetitions of the experiment, a given combination appears (recall that HF evaluations are four times as costly as LF ones). For instance, 10 out of 60 repetitions, the sequential design was composed of 20 LF evaluations and 4 HF evaluations.

where  $p_n(u) = P_n(\xi_2(u) \geq z^{\text{crit}})$  is the posterior mean of  $\mathbb{1}_{\xi_2(u) \geq z^{\text{crit}}}$ , and  $p_n(u) (1 - p_n(u))$  its posterior variance.

The experiment is repeated  $R = 60$  times—the simulator is deterministic, but randomness in the result comes from both the initial DoE and the use of a Monte Carlo procedure to sample from the posterior of the parameters. Figure 3a presents the evolution of the median estimation error, defined as  $\text{MedErr}_n = \text{median}_{1 \leq r \leq R} \|p_n^{(r)} - \mathbb{1}_{f_2 > z^{\text{crit}}}\|$  with  $\|\cdot\|$  the  $L^2$ -norm on  $\mathbb{U}$ , as function of the cost  $c_n = \sum_{i \leq n} C(\delta_i)$ . First, it appears clearly that high-fidelity evaluations are needed: the LF-SUR strategy achieves no significant error reduction with respect to the initial design. Second, we observe that the combination of low- and high-fidelity evaluations chosen by the MR-SUR strategy is more efficient, on average, than a purely high-fidelity sequential design. The actual number of evaluations on each level is summarized, for the 60 repetitions, in the table in Figure 3b: the MR-SUR strategy tends to use between two and five high-fidelity evaluations. (The recommendation of Xiong et al. (2013)—observing the low-fidelity level twice as many times as the high-fidelity one—would correspond here to six high-fidelity evaluations.)

### 4.3 Random damped harmonic oscillator

We now assess the performance of MR-SUR on an example proposed by Au and Beck (2001). We consider a random damped harmonic oscillator, whose displacement  $Y$  is the

Level $\delta$	1 s	0.5 s	1/3 s	0.25 s	0.2 s	1/6 s	0.1 s	0.05 s	0.02 s	0.01 s
Cost <sup>-1</sup>	32.7	24.8	19.9	16.7	14.3	12.6	8.4	4.6	2	1
Initial DoE	180	60	20	10	5	0	0	0	0	0

Table 1: Levels of fidelity considered the damped harmonic oscillator example (Section 4.3). The highest level of fidelity is  $\delta = 0.01$  s. Note that the initial design only spans the first five levels.

solution of the stochastic ordinary differential equation

$$\ddot{Y}(t) + 2\zeta\omega_0\dot{Y}(t) + \omega_0^2 Y(t) = W(t), \quad t \in [0, t_{\text{end}}], \quad \dot{Y}(0) = 0, \quad Y(0) = 0, \quad (14)$$

where  $\omega_0$  is the resonance frequency of the oscillator,  $\zeta$  is a damping coefficient,  $W$  is a Gaussian white noise and  $t_{\text{end}} = 30$  s. The solution of Equation (14) is approximated using an exponential Euler scheme with time step  $\delta > 0$  (more details can be found in the Supplementary Material): we denote by  $Y_k^{(\delta)}$  the resulting approximation of  $Y$  at time steps  $t_k = k\delta$ ,  $k \in \mathbb{N}$ ,  $k \leq K_\delta = \lfloor t_{\text{end}}/\delta \rfloor$ . We will be interested in the maximal log-displacement  $\max_{t \leq t_{\text{end}}} \log |Y(t)|$ , that we approximate by  $Z(\omega_0, \zeta, \delta) = \max_{k \leq K_\delta} \log(|Y_k^{(\delta)}|)$ .

We view the random variable  $Z(\omega_0, \zeta, \delta)$  as the response of a multi-fidelity stochastic simulator, where  $\delta$  controls the level of fidelity and  $\mathbf{u} = (\omega_0, \zeta) \in \mathbb{R}_+^2$  is the vector of input variables. In this problem, the QoI is the function  $Q$  defined by

$$Q(\omega_0, \zeta) = \mathbf{P}(Z(\omega_0, \zeta, \delta^{\text{ref}}) > z^{\text{crit}}), \quad 0 \leq \omega_0 \leq 30, \quad 0 \leq \zeta \leq 1,$$

where  $\delta^{\text{ref}} = 0.01$  s denotes the level of highest fidelity and  $z^{\text{crit}} = -3$  is a given critical threshold. The computational cost of  $Z$  is an affine function of  $1/\delta$ :  $C(\delta) = a/\delta + b$ . After normalization to have  $C(\delta^{\text{ref}}) = 1$ , the coefficients are  $a = 9.792 \times 10^{-3}$  and  $b = 2.08 \times 10^{-2}$ .

A good approximation of the output distributions is obtained if we assume  $Z(\omega_0, \zeta, \delta) | \xi, \lambda \sim \mathcal{N}(\xi(\omega_0, \zeta, \delta^{\text{ref}}), \lambda(\delta))$ , where the variance only depends on the fidelity level. This assumption makes it possible to write

$$Q(\omega_0, \zeta) = \Phi \left( \frac{\xi(\omega_0, \zeta, \delta^{\text{ref}}) - z^{\text{crit}}}{\sqrt{\lambda(\delta^{\text{ref}})}} \right).$$

The mean function  $\xi$  is modeled by the additive Gaussian process model (2)–(4) of Section 2.2, where the variance  $\lambda$  is log-Gaussian as in Section 2.3 and the prior distributions for the parameters are set as in Stroh et al. (2017b). The posterior mean  $\hat{Q}_n(\omega_0, \zeta) = \mathbf{E}_n(Q(\omega_0, \zeta))$  is used to estimate  $Q$ .

*Remark 6.* In this particular application, it was found that the value  $c = 0.99$  proposed by Stroh et al. (2017b), for the correlation between logarithms of noise variances, gave too much influence to the levels of lowest fidelity. The value  $c = 0.9$  was used instead.

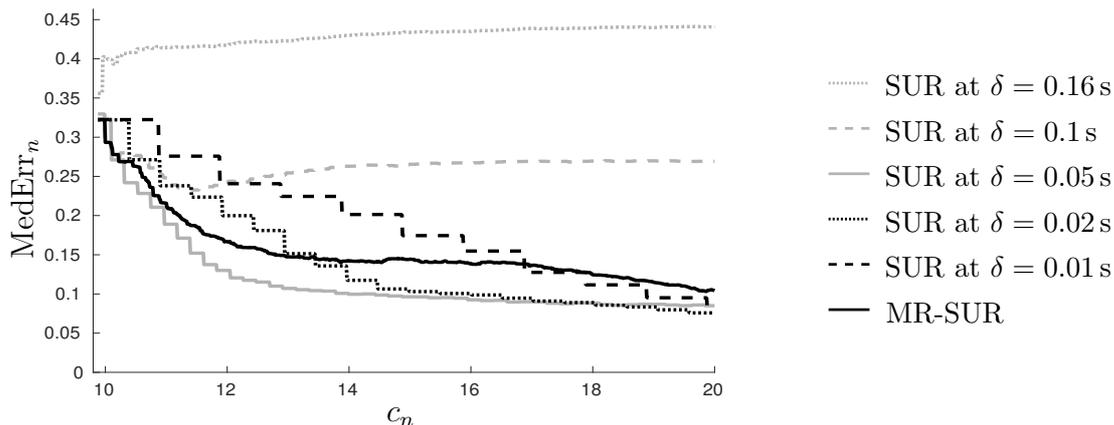


Figure 4: Median estimation error as a function of the cost for the oscillator test case.

In this example we consider  $S = 10$  levels, and the initial design is an NLHS on the five first levels. The different levels of fidelity, their costs, and the initial design are summarized in Table 1. The total cost of the initial design is 9.88. The total simulation budget, taking into account the initial budget, is set to 20. We also use a very high simulation budget to compute a reference value for  $Q$ , which will be used to assess the estimation error.

We compare the MR-SUR strategy, using the integrated posterior variance (10) as the uncertainty measure, to five different SUR strategies based on the same uncertainty measure. All of them start with the same multi-fidelity initial design (see Table 1). Each SUR strategy corresponds to sampling on only one of the five highest-fidelity levels, using this particular fidelity level both in the definition of the uncertainty measure (10) and for the prediction of the QoI. The experiment is repeated 200 times with different initial designs, and the strategies are compared using the median estimation error as in Section 4.2. A regular grid of size  $100 \times 100$  is used to discretize the integral in Equation (10) and the  $L^2$  norm of the error. The result is shown in Figure 4.

Comparing first MR-SUR with the fixed-level SUR strategy targetting the same level (namely,  $\delta = \delta^{\text{ref}} = 0.01$ ), it can be seen that MR-SUR makes a good use of the additional degree of freedom that multi-fidelity brings. Indeed, MR-SUR achieves superior performance for any budget between 10 and 16, and then performs similarly to the fixed-level strategy. Figure 5 shows some examples of the sequence of levels selected by MR-SUR : the design typically explores the level  $\delta = 0.1$  first, and then gradually moves to levels of higher fidelity.

Turning now to the results of the other fixed-level SUR strategies, it is interesting to see that the results depends very much on the fidelity level that is chosen. MR-SUR is by far superior to the fixed-level strategies at  $\delta = 0.16$  or  $\delta = 0.17$ , for any budget,

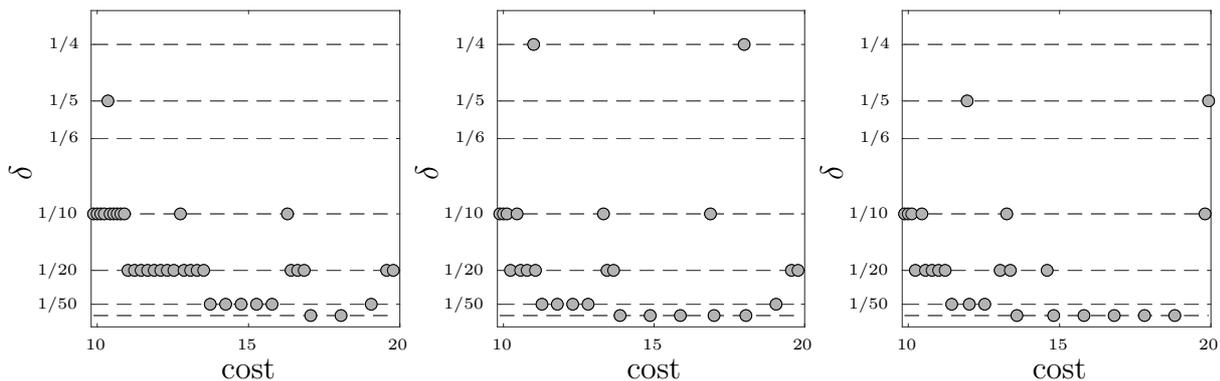


Figure 5: Three examples (out of 200) of the sequence of levels selected by the MR-SUR strategy for the oscillator test case.

but is outperformed by the fixed-level strategy at  $\delta = 0.05$  and, for budgets larger than approximately 14, also by the one at  $\delta = 0.02$ . The reason for this lies in two facts: First, the function  $(\omega_0, \zeta) \mapsto \mathbf{P}(Z(\omega_0, \zeta, \delta) > z^{\text{crit}})$  at these levels is close enough to the high-fidelity one. Second, the variance of the noise is cheaper to estimate at these levels. The MR-SUR strategy, *as implemented here*, fails to figure this out automatically since the uncertainty about the variance of the output is not taken into account.

#### 4.4 A fire safety example

In this section, we illustrate the MR-SUR strategy on a fire safety application. The goal is to assess the safety of a  $20 \text{ m} \times 12 \text{ m} \times 16 \text{ m}$  parallelepiped-shaped storage facility, with two  $2 \text{ m} \times 1 \text{ m}$  open doors and two  $2 \text{ m} \times 2 \text{ m}$  open windows. The propagation of smoke and heat is simulated using Fire Dynamics Simulator (FDS; see McGrattan et al., 2010), a state-of-the-art CFD software for fire engineering, which solves the transport equations using finite difference methods. The fire is located at the center of the room, and burns polyurethane. To assess fire safety, the values of several physical quantities are compared against regulatory thresholds—in this illustration, we focus on one of them only, called visibility and hereafter denoted by  $V$ . According to the ISO 13571 standard (2012), visibility must remain greater than  $z^{\text{crit}} = 5 \text{ m}$  to ensure safety during an evacuation.

Our FDS-based simulator will be treated as a stochastic simulator with nine input variables: three environmental variables (external temperature  $T_{\text{ext}}$ , atmospheric pressure  $P_{\text{atm}}$  and ambient temperature  $T_{\text{amb}}$ ) denoted by  $\mathbf{u}_e \in \mathbb{R}^3$ , five “scenario variables” (fire growth rate  $\alpha$ , fire area  $A_f$ , maximal heat release rate  $\dot{Q}''_h$ , total released energy per unit area  $q_{f_d}$  and soot yield  $Y_{\text{soot}}$ ) denoted by  $\mathbf{u}_s \in \mathbb{R}^5$ , and finally the size  $\delta$  of the spatial discretization mesh, which plays the role of a fidelity parameter. The reader is referred to Stroh et al.

Level $\delta$	50 cm	33 cm	25 cm	20 cm
Real cost	69 min	6 h	20 h	49 h
Normalized cost	1/42	1/8	1/2.5	1
Initial design	90	30	10	0

Table 2: The levels of fidelity on FDS.

(2017a) and Stroh (2018) for more details on the application.

*Remark 7.* Strictly speaking, the simulator considered in this section is actually deterministic, since the seed of the random number generator is fixed by the FDS software; cf. Stroh et al. (2017a) for details.

In this example, our objective is to estimate the probability that  $V$  becomes less than  $z^{\text{crit}}$  in a particular fire scenario, defined by  $\alpha = 0.1057 \text{ kW} \cdot \text{s}^{-2}$ ,  $A_f = 14 \text{ m}^2$ ,  $\dot{Q}_h'' = 460 \text{ kW} \cdot \text{m}^{-2}$ ,  $q_{f_d} = 450 \text{ MJ} \cdot \text{m}^{-2}$ , and  $Y_{\text{soot}} = 0.027 \text{ kg} \cdot \text{kg}^{-1}$ . The environmental inputs  $\mathbf{u}_e$  are assumed random and integrated according to an environmental distribution  $P_{\mathbf{u}_e}$ , which is a trivariate normal distribution with mean  $(10^\circ\text{C}, 100 \text{ kPa}, 22.5^\circ\text{C})$ , variances equal to  $(20/3^\circ\text{C}, 2/3 \text{ kPa}, 2.5^\circ\text{C})^2$ , and a correlation coefficient of 0.8 between the temperatures. The QoI is  $Q = \int_{\mathbf{U}_e} p(\mathbf{u}_e, \mathbf{u}_s, \delta^{\text{ref}}) dP_{\mathbf{u}_e}(\mathbf{u}_e)$ , where  $\delta^{\text{ref}} = 20 \text{ cm}$  is the reference level and  $p(\mathbf{u}_e, \mathbf{u}_s, \delta^{\text{ref}}) = \mathbf{P}(V < z^{\text{crit}} | \mathbf{u}_e, \mathbf{u}_s, \delta^{\text{ref}})$ .

Four levels of fidelity will be considered for running simulations: the reference level  $\delta^{\text{ref}} = 20 \text{ cm}$ , and three levels of lower fidelity ( $\delta = 50 \text{ cm}$ ,  $33 \text{ cm}$  and  $25 \text{ cm}$ ). Table 2 shows the correspondence between levels and computation times, assuming for simplicity that the computation time does not depend on the input variables (the fire area  $A_f$  and the total released energy per unit area  $q_{f_d}$ , which together control the amount  $A_f q_{f_d}$  of combustible material to be burnt, actually have some influence on the computation time, which is neglected). Four independent initial NLHS designs of size  $n = 130$ , distributed across the first three levels of fidelity as shown in Table 2, are available from previous studies. The normalized cost of each initial design is 9.89 (i.e., 20 days). A reference value for  $Q$  has also been obtained from 150 Monte Carlo simulations, distributed on the highest fidelity level using  $P_{\mathbf{u}_e}$ . This reference value has a normalized cost of 150 (i.e., about ten months).

We run the MR-SUR strategy starting from our four initial designs, using a supplementary budget of 24 for each run (about 48.7 days). The underlying Bayesian model is the same as in Section 4.3, the QoI  $Q$  is estimated using the posterior mean  $\hat{Q}_n = \mathbf{E}_n(Q)$ , and the measure of uncertainty is the integrated posterior variance

$$H_n = \int \text{var}_n(p(\mathbf{u}_e, \mathbf{u}_s, \delta^{\text{ref}})) dP_{\mathbf{u}_e}(\mathbf{u}_e),$$

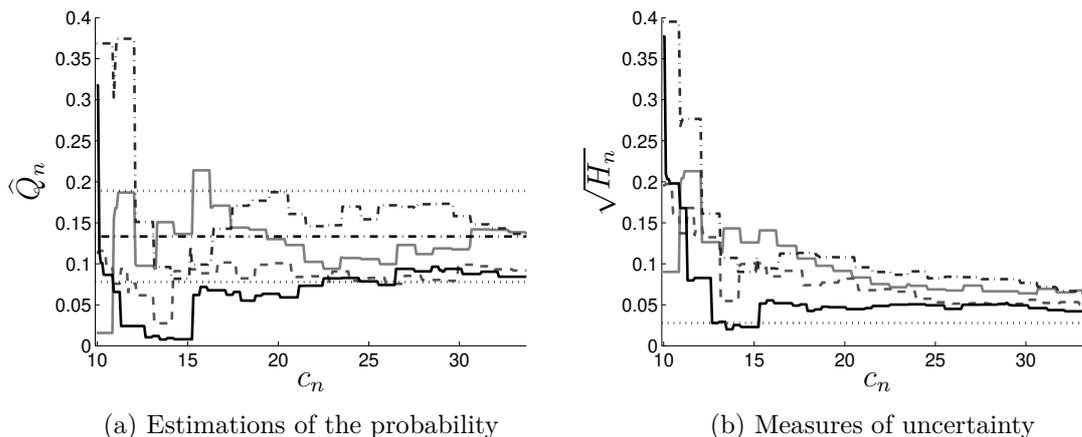


Figure 6: Result of four repetitions for the fire safety example. (a) Estimated probability as a function of the cost. The horizontal lines correspond to the Monte Carlo reference (dash-dotted line: mean; dotted lines: two-standard-deviation interval). (b) Square root of the measure of uncertainty (upper bound on the posterior standard deviation of the probability). The horizontal dotted line is the Monte Carlo standard deviation.

which is a special case of Equation (6). The corresponding SUR criterion is similar to (11), with an integral over the environmental variables only. The result is shown in Figure 6. We can see on Figure 6a that the estimations are initially, in three out of four cases, incompatible with the Monte Carlo one, but tend to get closer to the reference value when more simulations are carried out using the MR-SUR strategy. Figure 6b shows the measure of uncertainty as a function of the cost: the uncertainty is large at the beginning of the sequential design and rapidly becomes smaller, as expected, as the MR-SUR strategy proceeds. (Note that the cost of the whole design is approximately  $9.9 + 24 = 33.9$ , which is must cheaper than the cost of 150 of the Monte Carlo reference.)

## 5 Conclusion

The main contribution of this article is to unify and extend several methods of the literature of Bayesian sequential design of experiments for multi-fidelity numerical simulators. The unification that we propose is cast in the framework of Stepwise Uncertainty Reduction (SUR) strategies: when the accuracy of computer simulations can be chosen by the user, a natural extension of SUR strategies is to consider sampling criteria built as the ratio between the expected reduction of uncertainty and the simulation cost. We call this approach Maximal Rate of Stepwise Uncertainty Reduction (MR-SUR). It can be applied to deterministic or stochastic simulators. Our numerical experiments show that the MR-SUR

approach typically provides estimations which, for a given computational cost, are never much worse, and often better, than the best SUR strategy using a single level of fidelity.

Further work directions could be considered in the future. For instance, there is no explicit ingredient in MR-SUR strategies that tells the procedure to “learn” the model, and in particular, to learn the correlations between the levels of fidelity, and the variance of the output in the stochastic case. It seems to us that this would be important, particularly when simulations are very expensive and the simulation budget is very limited, as in our fire safety application. Using a fully Bayesian approach would somehow answer this problem, as the uncertainty about the model would be propagated to the uncertainty about the QoI.

Another useful extension would be to address the case where the cost function is not known beforehand. As already mentioned, when sufficient prior knowledge about this function is available, a few pilot runs at several levels of fidelity are typically used to estimate the cost. In situations where such knowledge is lacking, the Bayesian approach can in principle be extended to this function as well (as suggested by Swersky et al., 2013), which can then be learned on the fly. This idea, however, raises the question of how to adapt the MR-SUR strategy when the cost function is uncertain.

Finally, a third important research direction would be to address the design of parallel multi-fidelity simulations. As mentioned earlier in the article, leveraging batch-sequential SUR criteria makes it possible in principle to address the case of parallel batches of simulations with the same cost and run time—but the computational savings that can be obtained using such designs remain to be evaluated. Moreover, it seems even more promising to allow parallel simulations with different accuracies, costs and/or run times; but what would be a principled approach of (asynchronous) resource allocation in such a setting?

**Acknowledgment.** This work was supported by the French Association Nationale de la Recherche et de la Technologie (ANRT) under grant CIFRE 2014/1350.

## SUPPLEMENTARY MATERIAL

**Complementary information:** Generalization and proof of the SUR sampling criterion (11), a short bibliography about nonsequential multi-fidelity designs, and more detailed information regarding the numerical experiments. (PDF file)

*(In the published version of the article, this supplementary material is provided as a separate PDF file, available from the web site of the publisher. In this author-generated post-print version: see below.)*

**Source code:** Matlab programs to reproduce the numerical experiments of Sections 3.3, 4.2 and 4.3, including the modified version of the STK toolbox for Matlab/Octave (Bect et al., 2017) that was used for these experiments.

## References

- Au, S.-K. and Beck, J. L. (2001). Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4):263–277.
- Azzimonti, D., Ginsbourger, D., Chevalier, C., Bect, J., and Richet, Y. (2021). Adaptive design of experiments for conservative estimation of excursion sets. *Technometrics*, 63(1):13–26.
- Bect, J., Bachoc, F., and Ginsbourger, D. (2019). A supermartingale approach to Gaussian process based sequential design of experiments. *Bernoulli*, 25(4A):2883–2919.
- Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E. (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3):773–793.
- Bect, J., Vazquez, E., et al. (2017). STK: a Small (Matlab/Octave) Toolbox for Kriging. Release 2.5. <http://kriging.sourceforge.net>.
- Binois, M., Gramacy, R. B., and Ludkovski, M. (2018). Practical heteroscedastic Gaussian process modeling for large simulation experiments. *Journal of Computational and Graphical Statistics*, 27(4):808–821.
- Boukouvalas, A. and Cornford, D. (2009). Learning heteroscedastic Gaussian processes for complex datasets. Technical report, Neural Computing Research Group, Aston University.

- Brooks, C. J., Forrester, A. I. J., Keane, A. J., and Shahpar, S. (2011). Multi-fidelity design optimisation of a transonic compressor rotor. In *9<sup>th</sup> European Conf. Turbomachinery Fluid Dynamics and Thermodynamics*, Istanbul, Turkey.
- Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y. (2014). Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics*, 56(4):455–465.
- Chopin, N. (2002). A sequential particle filter method for static models. *Biometrika*, 89(3):539–552.
- Cochet, B., Jinaphanh, A., Heulers, L., and Jacquet, O. (2014). Capabilities overview of the MORET 5 Monte Carlo code. In *SNA+MC 2013 Joint International Conference on Supercomputing in Nuclear Applications + Monte Carlo*, page 12 pages.
- Demeyer, S., Fischer, N., and Marquis, D. (2017). Surrogate model based sequential sampling estimation of conformance probability for computationally expensive systems: application to fire safety science. *Journal de la Société Française de Statistique*, 158(1):111–138.
- Elsayed, K. (2015). Optimization of the cyclone separator geometry for minimum pressure drop using co-kriging. *Powder Technology*, 269:409–424.
- Feliot, P., Bect, J., and Vazquez, E. (2017). A Bayesian approach to constrained single- and multi-objective optimization. *Journal of Global Optimization*, 67(1):97–133.
- Forrester, A. I. J., Sóbester, A., and Keane, A. J. (2007). Multi-fidelity optimization via surrogate modelling. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 463(2088):3251–3269.
- Frazier, P. I., Powell, W. B., and Dayanik, S. (2009). The knowledge-gradient policy for correlated normal beliefs. *INFORMS Journal on Computing*, 21(4):599–613.
- Giselle Fernández-Godino, M., Park, C., Kim, N.-H., and Haftka, R. T. (2016). Review of multi-fidelity models. Technical report, University of Florida.
- Goh, J., Bingham, D., Holloway, J. P., Grosskopf, M. J., Kuranz, C. C., and Rutter, E. (2013). Prediction and computer model calibration using outputs from multifidelity simulators. *Technometrics*, 55(4):501–512.

- Goldberg, P. W., Williams, C. K. I., and Bishop, C. M. (1998). Regression with input-dependent noise: A Gaussian process treatment. In *Advances in neural information processing systems*, pages 493–499.
- Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242.
- Hage Hassan, M., Remy, G., Krebs, G., Marchand, C., Dupré, L., and Dr Guillaume Crevecoeur, P. (2014). Radial output space mapping for electromechanical systems design. *COMPEL - The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, 33(3):965–975.
- He, X., Tuo, R., and Wu, C. J. (2017). Optimization of multi-fidelity computer experiments via the EQIE criterion. *Technometrics*, 59(1):58–68.
- Huang, D., Allen, T. T., Notz, W. I., and Miller, R. A. (2006). Sequential kriging optimization using multiple-fidelity evaluations. *Structural and Multidisciplinary Optimization*, 32(5):369–382.
- ISO 13571 standard (2012). Life-threatening components of fire — Part 1: Guidelines for the estimation of time to compromised tenability in fires. Technical report, International Organization for Standardization (ISO), Geneva, Switzerland.
- Johnson, R. A. (1960). An information theory approach to diagnosis. *IRE Transactions on Reliability and Quality Control*, RQC-9(1):35–35.
- Jones, D. R., Schonlau, M., and Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492.
- Kansal, N. J. and Chana, I. (2012). Cloud load balancing techniques: A step towards green computing. *IJCSI International Journal of Computer Science Issues*, 9(1):238–246.
- Kennedy, M. C. and O’Hagan, A. (2000). Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13.
- Kersting, K., Plagemann, C., Pfaff, P., and Burgard, W. (2007). Most likely heteroscedastic Gaussian process regression. In *Proceedings of the 24<sup>th</sup> international conference on Machine learning (ICML’07)*, volume 227, pages 393–400.
- Koziel, S., Leifsson, L., Couckuyt, I., and Dhaene, T. (2013). Robust variable-fidelity optimization of microwave filters using co-kriging and trust regions. *Microwave and Optical Technology Letters*, 55(4):765–769.

- Kushner, H. J. (1964). A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *J. Basic Engineering*, 86:97–106.
- Kuya, Y., Takeda, K., Zhang, X., and Forrester, A. I. J. (2011). Multifidelity surrogate modeling of experimental and computational aerodynamic data sets. *AIAA journal*, 49(2):289–298.
- Le Gratiet, L. (2013). *Multi-fidelity Gaussian process regression for computer experiments*. PhD thesis, Université Paris-Diderot-Paris VII.
- Le Gratiet, L. and Cannamela, C. (2015). Cokriging-based sequential design strategies using fast cross-validation techniques for multi-fidelity computer codes. *Technometrics*, 57(3):418–427.
- Le Gratiet, L. and Garnier, J. (2014). Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification*, 4(5):365–386.
- Marques, A., Lam, R., and Willcox, K. (2018). Contour location via entropy reduction leveraging multiple information sources. In *Advances in neural information processing systems*, pages 5217–5227.
- Marques, A. N., Opgenoord, M. M. J., Lam, R. R., Chaudhuri, A., and Willcox, K. E. (2020). Multifidelity method for locating aeroelastic flutter boundaries. *AIAA Journal*, 58(4):1772–1784.
- Marrel, A., Iooss, B., Da Veiga, S., and Ribatet, M. (2012). Global sensitivity analysis of stochastic computer models with joint metamodels. *Statistics and Computing*, 22(3):833–847.
- McGrattan, K., McDermott, R., Hostikka, S., and Floyd, J. (2010). Fire dynamics simulator (version 5), user’s guide. NIST special publication 1019-5.
- Mockus, J., Tiesis, V., and Zilinskas, A. (1978). The application of Bayesian methods for seeking the extremum. In Dixon, L. and Szego, G., editors, *Towards Global Optimization*, volume 2, pages 117–129. Elsevier.
- Perdikaris, P. and Karniadakis, G. E. (2016). Model inversion via multi-fidelity Bayesian optimization: a new paradigm for parameter estimation in haemodynamics, and beyond. *Journal of The Royal Society Interface*, 13(118).

- Picheny, V. and Ginsbourger, D. (2013). A nonstationary space-time Gaussian process model for partially converged simulations. *SIAM/ASA Journal on Uncertainty Quantification*, 1(1):57–78.
- Picheny, V., Ginsbourger, D., Richet, Y., and Caplin, G. (2013). Quantile-based optimization of noisy computer experiments with tunable precision. *Technometrics*, 55(1):2–13.
- Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R. T., and Kim, N.-H. (2010). Adaptive designs of experiments for accurate approximation of a target region. *Journal of Mechanical Design*, 132(7):071008.
- Poloczek, M., Wang, J., and Frazier, P. (2017). Multi-information source optimization. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30, pages 4288–4298. Curran Associates, Inc.
- Qian, P. Z. G. (2009). Nested latin hypercube designs. *Biometrika*, 96(4):957–970.
- Qian, P. Z. G. and Wu, C. F. J. (2008). Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics*, 50(2):192–204.
- Qian, Z., Seepersad, C. C., Joseph, V. R., Allen, J. K., and Wu, C. F. J. (2006). Building surrogate models based on detailed and approximate simulations. *Journal of Mechanical Design*, 128(4):668–677.
- Ranjan, P., Bingham, D., and Michailidis, G. (2008). Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, 50(4):527–541.
- Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical science*, 4(4):409–423.
- Santner, T. J., Williams, B. J., and Notz, W. I. (2018). *The Design and Analysis of Computer Experiments. Second Edition*. Springer Series in Statistics. Springer, New York.
- Snoek, J., Larochelle, H., and Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems*, volume 25, pages 2951–2959. Curran Associates, Inc.
- Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics. Springer, New York.

- Stroh, R. (2018). *Sequential design of numerical experiments in multi-fidelity : Application to a fire simulator*. PhD thesis, Université Paris-Saclay.
- Stroh, R., Bect, J., Demeyer, S., Fischer, N., Marquis, D., and Vazquez, E. (2017a). Assessing fire safety using complex numerical models with a Bayesian multi-fidelity approach. *Fire Safety Journal*, 91:1016–1025.
- Stroh, R., Bect, J., Demeyer, S., Fischer, N., and Vazquez, E. (2017b). Integrating hyperparameter uncertainties in a multi-fidelity Bayesian model for the estimation of a probability of failure. In 11<sup>th</sup> *Advanced Mathematical and Computational Tools in Metrology and Testing Conference (AMCTM 2017)*, Glasgow, Scotland.
- Swersky, K., Snoek, J., and Adams, R. P. (2013). Multi-task Bayesian optimization. In Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems*, volume 26, pages 2004–2012. Curran Associates, Inc.
- Tuo, R., Wu, C. F. J., and Yu, D. (2014). Surrogate modeling of computer experiments with different mesh densities. *Technometrics*, 56(3):372–380.
- Villemonteix, J., Vazquez, E., and Walter, E. (2009). An informational approach to the global optimization of expensive-to-evaluate functions. *Journal of Global Optimization*, 44:509–534.
- Xiong, S., Qian, P. Z. G., and Wu, C. F. J. (2013). Sequential design and analysis of high-accuracy and low-accuracy computer codes. *Technometrics*, 55(1):37–46.
- Yi, J., Wu, F., Zhou, Q., Cheng, Y., Ling, H., and Liu, J. (2021). An active-learning method based on multi-fidelity kriging model for structural reliability analysis. *Structural and Multidisciplinary Optimization*, 63:1–23.

## SUPPLEMENTARY MATERIAL

### SM-1 Introduction

This document contains supplementary material for the article “Sequential design of multi-fidelity computer experiments: maximizing the rate of stepwise uncertainty reduction”. It is organized as follows. Section SM-2 provides a short literature review on nonsequential designs for multi-fidelity, which complements the literature review on sequential designs given in Section 3.1 of the article. Section SM-3 provides (with proof) a new SUR criterion, which covers as a special case the criterion provided (without proof) in Section 3.2 of the article. Finally, Section SM-4 provides additional information regarding the examples presented in Sections 4.2 and 4.3 of the article.

### SM-2 Nonsequential designs in multi-fidelity

In this section, we provide a very brief literature review on nonsequential designs for multi-fidelity.

A common recommendation for multi-level designs is nesting. A multi-level design is nested when any observed point at a level  $\delta^{(s)}$  is also observed at every lower-fidelity level  $\delta^{(s')}$ ,  $s' < s$ . Furthermore, space-filling designs are also expected to ensure observations in the whole input domain, as usual in Gaussian process regression.

A simple method to create a nested design is proposed by Forrester et al. (2007). It draws a maximin Latin Hypercube Sampling (LHS) at the lowest fidelity design, and then selects subsets of this LHS at the next levels. Le Gratiet and Garnier (2014) suggest to start with the highest-fidelity level and add points on the lower-fidelity levels to ensure better space-filling properties. The reader is also referred to Rennen et al. (2010) for a method which applies when there is only two levels of fidelity.

In our work, we use the method proposed by Qian (2009), which construct Nested Latin Hypercube Sampling (NLHS). An NLHS is a nested design with the property of being an LHS at each level of fidelity. We add a maximin optimization step to ensure that the design is space-filling at each level.

Note that this method was extended in several directions. He and Qian (2011), Yang et al. (2014), Guo et al. (2017), and Xu et al. (2017) propose methods to build NLHS with particular structures, such as orthogonality.

## SM-3 A new SUR criterion

### SM-3.1 Criterion definition and result statement

Let  $\xi = (\xi(\mathbf{x}))_{\mathbf{x} \in \mathbb{X}}$  denote a Gaussian process prior for the mean response of a stochastic simulator with Gaussian responses—i.e., a simulator which produces random responses

$$Z \mid \xi \sim \mathcal{N}(\xi(\mathbf{x}), \lambda(\mathbf{x})),$$

where  $\mathbf{x} \in \mathbb{X}$  denotes the vector of inputs of the simulator and  $\lambda : \mathbb{X} \rightarrow [0, +\infty)$  is a known variance function. Assume that the quantity of interest is the probability function  $\alpha : \mathbb{X} \rightarrow [0, 1]$  defined by

$$\alpha(\mathbf{x}) = \mathbb{P}(Z_{\mathbf{x}} > z^{\text{crit}}),$$

where  $z^{\text{crit}} \in \mathbb{R}$  is a given threshold and  $Z_{\mathbf{x}}$  denotes a (future) response of the simulator with  $\mathbf{x}$  as the input. More explicitly, we have

$$\alpha(\mathbf{x}) = \Phi\left(\frac{\xi(\mathbf{x}) - z^{\text{crit}}}{\sqrt{\lambda(\mathbf{x})}}\right). \quad (\text{SM1})$$

Let  $\hat{\alpha}_n(\mathbf{x})$  denote the posterior mean of  $\alpha(\mathbf{x})$  given  $n$  responses  $Z_1, \dots, Z_n$  of the simulator at design points  $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{X}$  (possibly selected in a sequential manner). Let  $\mu$  denote a positive, bounded measure on  $\mathbb{X}$ , and consider the measure of uncertainty  $H_n$  defined by

$$H_n = \mathbb{E}_n\left(\int (\alpha(\mathbf{x}) - \hat{\alpha}_n(\mathbf{x}))^2 \mu(d\mathbf{x})\right) = \int \text{var}_n(\alpha(\mathbf{x})) \mu(d\mathbf{x}).$$

The following result provides a tractable expression for the corresponding SUR criterion for a batch  $\tilde{\mathbf{x}} = (\tilde{\mathbf{x}}_l)_{1 \leq l \leq q} \in \mathbb{X}^q$  of candidate points. The criterion presented in the article corresponds to the fully sequential case ( $q = 1$ ) and to a particular choice of the measure  $\mu$ .

**Proposition 1.** *Let  $m_n$  (resp.  $k_n$ ) denote the posterior mean (resp. the posterior covariance) of  $\xi$  given  $n$  observations. Let*

$$J_n(\tilde{\mathbf{x}}) = \mathbb{E}\left(H_{n+q} \mid \mathbf{X}_{n+1} = \tilde{\mathbf{x}}_1, \dots, \mathbf{X}_{n+q} = \tilde{\mathbf{x}}_q\right)$$

and  $G_n(\tilde{\mathbf{x}}) = H_n - J_n(\tilde{\mathbf{x}})$ . Then,

$$J_n(\tilde{\mathbf{x}}) = \int \left[ \Phi_2(u_n(\mathbf{x}), a_n(\mathbf{x}); r_n(\mathbf{x})) - \Phi_2(a_n(\mathbf{x}), a_n(\mathbf{x}); \tilde{r}_n(\mathbf{x}; \tilde{\mathbf{x}})) \right] \mu(d\mathbf{x})$$

and

$$G_n(\tilde{\mathbf{x}}) = \int \left[ \Phi_2(a_n(\mathbf{x}), a_n(\mathbf{x}); \tilde{r}_n(\mathbf{x}; \tilde{\mathbf{x}})) - \Phi(a_n(\mathbf{x}))^2 \right] \mu(d\mathbf{x}),$$

with  $\Phi$  the cdf of the standard normal distribution,  $\Phi_2(\cdot, \cdot; \rho)$  the cdf of the standard bivariate normal distribution with correlation  $\rho$ , and

$$\begin{aligned} a_n(\mathbf{x}) &= (m_n(\mathbf{x}) - z^{\text{crit}}) / \sqrt{v_n(\mathbf{x})}, & v_n(\mathbf{x}) &= k_n(\mathbf{x}, \mathbf{x}) + \lambda(\mathbf{x}), \\ r_n(\mathbf{x}) &= k_n(\mathbf{x}, \mathbf{x}) / v_n(\mathbf{x}), & \tilde{r}_n(\mathbf{x}; \tilde{\mathbf{x}}) &= \nu_n(\mathbf{x}, \mathbf{x}; \tilde{\mathbf{x}}) / v_n(\mathbf{x}), \\ \nu_n(\mathbf{x}, \mathbf{x}'; \tilde{\mathbf{x}}) &= k_n(\mathbf{x}, \mathbf{x}') - k_{n+q}(\mathbf{x}, \mathbf{x}'; \tilde{\mathbf{x}}) = k_n(\tilde{\mathbf{x}}, \mathbf{x})^T K_n(\tilde{\mathbf{x}}, \tilde{\mathbf{x}})^{-1} k_n(\tilde{\mathbf{x}}, \mathbf{x}'), \\ k_n(\tilde{\mathbf{x}}, \mathbf{x}) &= (k_n(\tilde{\mathbf{x}}_l, \mathbf{x}))_{1 \leq l \leq q}, & K_n(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}) &= (k_n(\tilde{\mathbf{x}}_l, \tilde{\mathbf{x}}_{l'}) + \lambda(\tilde{\mathbf{x}}_l) \cdot \delta_{l=l'})_{1 \leq l, l' \leq q}. \end{aligned}$$

*Remark SM1.* In the expressions of  $J_n(\tilde{\mathbf{x}})$  and  $G_n(\tilde{\mathbf{x}})$ , the only part which depends on the future design  $\tilde{\mathbf{x}}$  is  $\int \Phi_2(a_n(\mathbf{x}), a_n(\mathbf{x}); \tilde{r}_n(\mathbf{x}; \tilde{\mathbf{x}})) \mu(d\mathbf{x})$ , which must be maximized.

### SM-3.2 A useful identity

Let  $\tilde{\Phi}_d(\cdot; m, K)$  denote the cumulative distribution function of the  $d$ -variate normal distribution with mean  $m$  and covariance matrix  $K$ . The following identity is used by [Chevalier et al. \(2014\)](#) for the computation of SUR criteria similar to ours.

**Lemma 2.** *Let  $W \sim \mathcal{N}(m, K)$  be a  $d$ -dimensional Gaussian vector. Then, for any mean vector  $m'$  and covariance matrix  $K'$ ,*

$$\mathbb{E} \left( \tilde{\Phi}_d(W; m', K') \right) = \tilde{\Phi}_d(m; m', K + K').$$

*Proof.* Let  $W' \sim \mathcal{N}(m', K')$  be independent from  $W$ . Then

$$\mathbb{E} \left( \tilde{\Phi}_d(W, m', K') \right) = \mathbb{E} \left( \mathbb{P} \left( W' \leq W \mid W \right) \right) = \mathbb{P}(W' \leq W),$$

and, using that  $W' - W = W'' - m$  with  $W'' \sim \mathcal{N}(m', K + K')$ ,

$$\mathbb{P}(W' \leq W) = \mathbb{P}(W'' \leq m) = \tilde{\Phi}_d(m; m', K + K'). \quad \square$$

**Corollary 3.** *Let  $m, m' \in \mathbb{R}$ ,  $v, v' \in [0, +\infty)$  and  $W \sim \mathcal{N}(m, v)$ . Then*

$$\mathbb{E} \left( \Phi \left( \frac{W - m'}{\sqrt{v'}} \right) \right) = \Phi \left( \frac{m - m'}{\sqrt{v + v'}} \right)$$

and

$$\mathbb{E} \left( \Phi \left( \frac{W - m'}{\sqrt{v'}} \right)^2 \right) = \Phi_2 \left( \frac{m - m'}{\sqrt{v + v'}}, \frac{m - m'}{\sqrt{v + v'}}; \frac{v}{v + v'} \right).$$

### SM-3.3 Proof of Proposition 1

*Proof.* Recall from Equation (SM1) that

$$\alpha(\mathbf{x}) = \Phi\left(\frac{\xi(\mathbf{x}) - z^{\text{crit}}}{\sqrt{\lambda(\mathbf{x})}}\right).$$

It follows from Corollary 3 that

$$\mathbb{E}_{n+q}(\alpha(\mathbf{x})) = \Phi\left(\frac{m_{n+q}(\mathbf{x}) - z^{\text{crit}}}{\sqrt{\lambda(\mathbf{x}) + k_{n+q}(\mathbf{x}, \mathbf{x})}}\right) = \Phi\left(\frac{m_{n+q}(\mathbf{x}) - z^{\text{crit}}}{\sqrt{v_{n+q}(\mathbf{x})}}\right), \quad (\text{SM2})$$

and thus

$$\text{var}_{n+q}(\alpha(\mathbf{x})) = \mathbb{E}_{n+q}(\alpha(\mathbf{x})^2) - \Phi\left(\frac{m_{n+q}(\mathbf{x}) - z^{\text{crit}}}{\sqrt{v_{n+q}(\mathbf{x})}}\right)^2. \quad (\text{SM3})$$

Let us now compute separately the expectation with respect to  $\mathbb{P}_n$  of the two terms in the right-hand side of Equation (SM3). For the first term we have

$$\begin{aligned} \mathbb{E}_n(\mathbb{E}_{n+q}(\alpha(\mathbf{x})^2)) &= \mathbb{E}_n(\alpha(\mathbf{x})^2) = \mathbb{E}_n\left(\Phi\left(\frac{\xi(\mathbf{x}) - z^{\text{crit}}}{\sqrt{\lambda(\mathbf{x})}}\right)^2\right) \\ &= \Phi_2(a_n(\mathbf{x}), a_n(\mathbf{x}); r_n(\mathbf{x})), \end{aligned} \quad (\text{SM4})$$

where we have applied the second part of Corollary 3 with  $m = m_n(\mathbf{x})$ ,  $v = k_n(\mathbf{x}, \mathbf{x})$ ,  $m' = z^{\text{crit}}$ , and  $v' = \lambda(\mathbf{x})$ . For the second term we observe that, under  $\mathbb{P}_n$ ,  $m_{n+q}$  is a Gaussian process with mean  $m_n$  and covariance function  $\nu_n(\cdot, \cdot; \tilde{\mathbf{x}})$ . Therefore  $m_{n+q}(\mathbf{x}) \sim \mathcal{N}(m_n(\mathbf{x}), \nu(\mathbf{x}, \mathbf{x}; \tilde{\mathbf{x}}))$ , and it follows that

$$\mathbb{E}_n\left(\Phi\left(\frac{m_{n+q}(\mathbf{x}) - z^{\text{crit}}}{\sqrt{v_{n+q}(\mathbf{x})}}\right)^2\right) = \Phi_2(a_n(\mathbf{x}), a_n(\mathbf{x}); \tilde{r}_n(\mathbf{x})), \quad (\text{SM5})$$

where we have used again the second part of Corollary 3 with  $m = m_n(\mathbf{x})$ ,  $v = \nu_n(\mathbf{x}, \mathbf{x}; \tilde{\mathbf{x}})$ ,  $m' = z^{\text{crit}}$  and  $v' = v_{n+q}(\mathbf{x})$ . Indeed,

$$\begin{aligned} v + v' &= \nu_n(\mathbf{x}, \mathbf{x}; \tilde{\mathbf{x}}) + v_{n+q}(\mathbf{x}) \\ &= (k_n(\mathbf{x}, \mathbf{x}) - k_{n+q}(\mathbf{x}, \mathbf{x})) + (k_{n+q}(\mathbf{x}, \mathbf{x}) + \lambda(\mathbf{x})) = v_n(\mathbf{x}), \end{aligned}$$

therefore

$$\begin{aligned} \frac{m - m'}{\sqrt{v + v'}} &= \frac{m_n(\mathbf{x}) - z^{\text{crit}}}{\sqrt{v_n}} = a_n(\mathbf{x}), \\ \frac{v}{v + v'} &= \frac{\nu_n(\mathbf{x}, \mathbf{x}; \tilde{\mathbf{x}})}{v_n(\mathbf{x})} = \tilde{r}_n(\mathbf{x}). \end{aligned}$$

Combining Equations (SM3)–(SM5) and integrating on  $\mathbb{X}$  with respect to  $\mu$  yields the desired expression for  $J_n(\mathbf{x})$ . Similarly, combining Equation (SM2) with  $q = 0$  and Equation (SM4) we have

$$H_n = \int \left( \Phi_2(a_n(\mathbf{x}), a_n(\mathbf{x}); r_n(\mathbf{x})) - \Phi(a_n(\mathbf{x}))^2 \right) \mu(d\mathbf{x})$$

and the expression of  $G_n(\tilde{\mathbf{x}})$  follows. □

## SM-4 Additional information regarding the examples

### SM-4.1 One-dimensional example

This section provides additional information regarding the Bayesian model used in the “One-dimensional example” presented in Section 4.2 of the article.

The model used in this example is the one proposed by Kennedy and O’Hagan (2000) and reviewed in Section 2.1 of the article, with  $S = 2$  levels. The two independent Gaussian processes  $\eta_1$  and  $\eta_2$  are stationary processes with unknown constant means and Matérn covariance functions with regularity  $5/2$ :

$$\eta_s \sim \text{GP} \left( m_s, \sigma_s^2 \mathcal{M}_{5/2}(a_s(\cdot - \cdot)) \right), \quad s \in \{1, 2\}.$$

Independent priors are used for all the remaining parameters of the model:

- Improper uniform prior distributions on  $\mathbb{R}$  are used for the means  $m_s$ .
- The parameters of the covariance function follow log-normal distributions:

$$\begin{aligned} \log(\sigma_s^2) &\sim \mathcal{N}(2 \log(0.2), \log(100)^2), \\ \log(a_s) &\sim \mathcal{N}(\log(2), \log(10)^2). \end{aligned}$$

- Finally, the regression term between the two levels follows a normal prior distribution:

$$\rho \sim \mathcal{N}(1, 2^2).$$

### SM-4.2 Random damped harmonic oscillator

This section provides additional information regarding the explicit exponential Euler scheme used in the “Random damped harmonic oscillator” example (Section 4.3 of the article).

Consider a stochastic differential equation

$$d\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_t dt + \mathbf{b} dB_t,$$

where  $(\mathbf{Y}_t)$  is a vector-valued stochastic process,  $(B_t)$  a Wiener process,  $\mathbf{b}$  a vector and  $\mathbf{A}$  a matrix. Let  $\delta$  a time step, we would like to approximate  $(\mathbf{Y}_t)$  by a finite difference method:  $\mathbf{Y}_k^{(\delta)} \approx \mathbf{Y}_{k\delta}$ . The explicit exponential Euler scheme is a finite difference method proposed by Jentzen and Kloeden (2009) to ensure stability when approximating a stochastic equation. The method is to apply recursively the formula

$$\mathbf{Y}_{k+1}^{(\delta)} = \exp(\mathbf{A}\delta) \left[ \mathbf{Y}_k^{(\delta)} + \sqrt{2\pi S\delta} \cdot \mathbf{b} \cdot U \right]$$

with  $S$  the spectral intensity of the Brownian motion, and  $U$  a normal random vector.

In particular, for the application of Section 4.2, we have

$$d \begin{pmatrix} Y_t \\ \dot{Y}_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\omega_0^2 & -2\zeta\omega_0 \end{pmatrix} \begin{pmatrix} Y_t \\ \dot{Y}_t \end{pmatrix} dt + \begin{pmatrix} 0 \\ 1 \end{pmatrix} dB_t.$$

Consequently, the approximation with a finite time-step  $\delta$  is

$$\begin{pmatrix} Y_{k+1}^{(\delta)} \\ \dot{Y}_{k+1}^{(\delta)} \end{pmatrix} = \exp \begin{pmatrix} 0 & \delta \\ -\omega_0^2\delta & -2\zeta\omega_0\delta \end{pmatrix} \left[ \begin{pmatrix} Y_k^{(\delta)} \\ \dot{Y}_k^{(\delta)} \end{pmatrix} + \begin{pmatrix} 0 \\ \sqrt{2\pi S\delta}u \end{pmatrix} \right], \quad u \sim \mathcal{N}(0, 1).$$

## References for the Supplementary Material

- Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y. (2014). Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics*, 56(4):455–465.
- Forrester, A. I. J., Sóbester, A., and Keane, A. J. (2007). Multi-fidelity optimization via surrogate modelling. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 463(2088):3251–3269.
- Guo, B., Chen, X.-P., and Liu, M.-Q. (2017). Construction of latin hypercube designs with nested and sliced structures. *Statistical Papers*, pages 1–14.
- He, X. and Qian, P. Z. G. (2011). Nested orthogonal array-based latin hypercube designs. *Biometrika*, 98(3):721–731.
- Jentzen, A. and Kloeden, P. E. (2009). Overcoming the order barrier in the numerical approximation of stochastic partial differential equations with additive space–time noise. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 465(2102):649–667.
- Le Gratiet, L. and Garnier, J. (2014). Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification*, 4(5):365–386.

- Qian, P. Z. G. (2009). Nested latin hypercube designs. *Biometrika*, 96(4):957–970.
- Rennen, G., Husslage, B., Van Dam, E. R., and Den Hertog, D. (2010). Nested maximin latin hypercube designs. *Structural and Multidisciplinary Optimization*, 41(3):371–395.
- Xu, J., Duan, X., Wang, Z., and Yan, L. (2017). A general construction for nested latin hypercube designs. *Statistics & Probability Letters*.
- Yang, J., Liu, M.-Q., and Lin, D. K. J. (2014). Construction of nested orthogonal latin hypercube designs. *Statistica Sinica*, 24(1):211–219.