



**HAL**  
open science

# Optimization of Noisy Computer Experiments with Tunable Precision

Victor Picheny, David Ginsbourger, Yann Richet, Grégory Caplin

► **To cite this version:**

Victor Picheny, David Ginsbourger, Yann Richet, Grégory Caplin. Optimization of Noisy Computer Experiments with Tunable Precision. 2011. hal-00578550v1

**HAL Id: hal-00578550**

**<https://hal.science/hal-00578550v1>**

Preprint submitted on 25 Mar 2011 (v1), last revised 19 Mar 2012 (v3)

**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.

# Optimization of Noisy Computer Experiments with Tunable Precision

Victor Picheny

Ecole Centrale de Paris, Chatenay-Malabry, France

email: `victor.picheny@ecp.fr`

David Ginsbourger

University of Bern, Switzerland

email: `david.ginbourger@stat.unibe.ch`

Yann Richet

Institut de Radioprotection et de Sûreté Nucléaire, Fontenay-aux-Roses, France

email: `yann.richet@irsn.fr`

Gregory Caplin

Institut de Radioprotection et de Sûreté Nucléaire, Fontenay-aux-Roses, France

email: `gregory.caplin@irsn.fr`

February 7, 2011

## Abstract

This article addresses the issue of kriging-based optimization of stochastic simulators. Many of these simulators depends on factors that tune the level of precision of the response, the gain in accuracy being at a price of computational time. The contribution of this work is two-fold: firstly, we propose a quantile-based criterion for the sequential choice of experiments, in the fashion of the classical Expected Improvement criterion, which allows a rigorous treatment of heterogeneous response precisions. Secondly, we present a procedure that allocates on-line the computational time given to each measurement, allowing a better distribution of the computational effort and increased efficiency. Finally, the optimization method is applied to an original application in nuclear criticality safety.

**Keywords:** Noisy optimization, Kriging, Tunable fidelity.

## 1. INTRODUCTION

Using metamodels for facilitating optimization and statistical analysis of computationally expensive simulators has become commonplace. In particular, the kriging-based EGO algorithm (Jones, Schonlau and Welch 1998) and its underlying expected improvement (EI) criterion have been recognized as efficient tools for deterministic black-box optimization.

The way a simulator response follows the function of interest is called *fidelity*. Oftentimes, a large range of response fidelities is available by tuning factors that control the complexity of numerical methods. For instance, the precision of a finite element analysis can be controlled by the discretization technique or the solver convergence. When the response stems from Monte Carlo methods (which is often referred to as *stochastic simulators*), the accuracy (measured by response variance) is inversely proportional to sample size.

Such simulators are often called *noisy simulators*, since they return approximate solutions that depart from the exact value by an error term that can be considered as a random quantity. Optimization in this context raises critical issues. Having noise in the responses requires a proper adaptation of criteria and algorithms. Furthermore, for each simulation run, the user has to set a trade-off between computational cost and response precision. The choice of this trade-off greatly impacts the efficiency of the optimization.

Using metamodels for noisy optimization has been already addressed by several authors. Many approaches consider only two fidelity levels, and the low-fidelity model is used as a helping tool to choose the high-fidelity evaluations (Alexandrov, Lewis, Gumbert, Green and Newman 2000) (Gano, Renaud, Martin and Simpson 2006). In that case, metamodels are used to estimate the difference between simulators. More integrated approaches have also been proposed, based on modifications of the EGO algorithm. Huang, Allen, Notz and Miller (2006) proposed a criterion for hierarchical kriging models with finitely many levels of fidelity, that chooses at the same time the observation point and the fidelity. Forrester, Keane and Bressloff (2006) proposed a re-interpolation technique to filter out the noise, allowing the use of the standard EGO algorithm.

This work proposes two contributions to this framework. First, we define an extension of EI based on quantiles that enables a rigorous treatment of both continuous or discrete fidelities. The proposed criterion not only depends on the noise variances from the past, but also on the fidelity of the new candidate measurement. Hence, this criterion allows to choose both an input space point and a fidelity level at each iteration. Second, we study a procedure taking advantage of this additional degree of freedom. Once an input space point has been selected, computation time is invested on it until a stopping criterion is met. One of the advantages of such procedure is that it prevents from allocating too much time to poor designs, and allows spending more credit on the best ones.

In the next section, we describe the classical kriging-based optimization procedure, and its limitation with noisy functions. Then, the quantile-based EI criterion is proposed, followed by the on-line allocation pro-

cedure. Finally, an original application in nuclear criticality safety is implemented in the Promethee workbench and applied to the Monte Carlo criticality simulator MORET5 (Fernex, Heulers, Jacquet, Miss and Richet 2005).

## 2. NOTATIONS AND CONCEPTS

### 2.1 The noisy optimization problem

We consider a single objective, unconstrained optimization problem. The deterministic objective function  $y : \mathbf{x} \in D \subset \mathbb{R}^d \rightarrow y(\mathbf{x}) \in \mathbb{R}$  is here observed in noise. For a measurement at some  $\mathbf{x} \in D$ , the user doesn't have access to the exact  $y(\mathbf{x})$ , but to an approximate response  $y(\mathbf{x}) + \epsilon$ .  $\epsilon$  is assumed to be one realization of a "noise" random variable  $\epsilon$ , whose probability distribution may depend on  $\mathbf{x}$  and other variables, and which realizations might differ for different measurements of  $y$  at the same  $\mathbf{x}$ . So instead of referring to the measurements of  $y$  in terms of  $\mathbf{x}$ 's, we will denote by  $\tilde{y}_i = y(\mathbf{x}^i) + \epsilon_i$  the noisy measurements, where the  $\mathbf{x}^i$ 's are not necessarily all distinct.

### 2.2 Noise in computer experiments

In classical experiments, noise usually accounts for a large number of uncontrolled variables (variations of the experimental setup, measurement precision, etc.). In computer experiments, noise can have many sources, including modeling and discretization error, incomplete convergence, and finite sample size for Monte-Carlo methods (see for instance Forrester, Keane and Bressloff (2006) for a detailed discussion). Also, any deterministic simulator including some random input can be viewed as a stochastic one, considering that the objective function is a statistic of the output (typically, a mean or quantile) over the random inputs.

The nature of the noise depends on the associated simulator. When Monte-Carlo methods are involved in the output evaluation, error is independent from one run to each other, even for measurements with the same input variables. Such simulators are often referred to as *stochastic*, and are the main target for the method presented here.

In the framework of *multi-fidelity evaluations*, error is due to a simplification of the physics equations, geometry, or discretization (e.g. meshing in Finite Elements models). In that case, errors are likely to be strongly correlated, especially for simulations with similar fidelities, and repeated experiments provide the same observations. This situation has been addressed in the litterature (see Kennedy and O'Hagan (2000) and Gano et al. (2006) for modeling, Alexandrov et al. (2000) and Huang, Allen, Notz and Zeng (2006) for optimization) and will not be considered here.

When error is due to incomplete convergence, errors are also likely to be correlated. In the problem considered in Forrester, Bressloff and Keane (2006), simulations across the design space tend to converge in unison (errors are almost equal for two measurements with the same convergence level), which makes the partial convergence equivalent to a multi-fidelity problem. However, when the output convergence behaviour

varies substantially across the design space, the hypothesis of independency of the error between runs may become reasonable, especially if experiments are well spread in the design space and different convergence levels are used.

In the rest of this article, we stick to the stochastic simulation framework, and make the assumption that the observation noises are normally distributed, centered and independent from one run to each other:

$$\varepsilon_i \sim \mathcal{N}(0, \tau_i^2) \text{ independently.} \quad (1)$$

### 2.3 Experiments with tunable precision

As mentioned in the introduction, the precision of such simulators can be tuned by the user, for instance by changing the number of solver steps for incomplete convergence or the sample size for Monte-Carlo methods. Of course, the precision level increases with computational time.

Hence, we consider that for every measurement  $i$  ( $1 \leq i \leq n$ ), the noise variance  $\tau_i^2$  is controllable and decreases monotonically towards zero with the allocated computational time  $t_i$ . Then, the actual (inaccessible) objective function  $y$  is the response given by the simulator with an infinite computational time allocated at every  $\mathbf{x} \in D$ . The difference between the simulated and actual phenomena is not considered here.

It is assumed in this work that the noise variance is a known decreasing function of computation time:

$$\tau^2 : t \in [0, +\infty[ \longrightarrow \tau^2(t) \in [0, +\infty[ \quad (2)$$

Although some stochastic simulators, such as the one described in section 7.2, directly provide an accurate estimate of the output uncertainty, in most real applications a learning study is necessary, typically assuming a (simple) parametric form for the variance. In the case of Monte-Carlo simulators (or repeated experiments) and assuming small variations of the output across the design space, we have  $\tau^2(t) = C/t$ , where  $C$  is an unknown constant which can be estimated when building the kriging model, as described in the next section.

Finally, for simulators relying on Monte Carlo or on iterative solvers, the response corresponding to a given precision is not obtained directly but more as a limit of intermediate responses of lower precisions. For each measurement, the noisy response  $\tilde{y}_i$  is thus obtained as last term of a sequence of measurements  $\tilde{y}_i[1], \dots, \tilde{y}_i[b_i]$  with decreasing noise variances,  $\tau_i^2[1] > \dots > \tau_i^2[b_i]$ , where  $b_i \in \mathbb{N}$  is the number of calculation steps at the  $i^{\text{th}}$  measurement. Furthermore, each step is assumed here to correspond to one elementary computation time  $t_e \in ]0, +\infty[$ , so that  $\forall j \in \{1, \dots, b_i\}$ ,  $\tau_i^2[j] = \tau^2(j \times t_e)$ .

Figure 1 represents the convergence of the output of the stochastic simulator of section 7.2 for its nominal design values. Here, the variance is known accurately, and depicted by the 95% confidence interval. The curve *ytilde* represents the sequence  $\tilde{y}[j], j = 1 \dots 100$ .

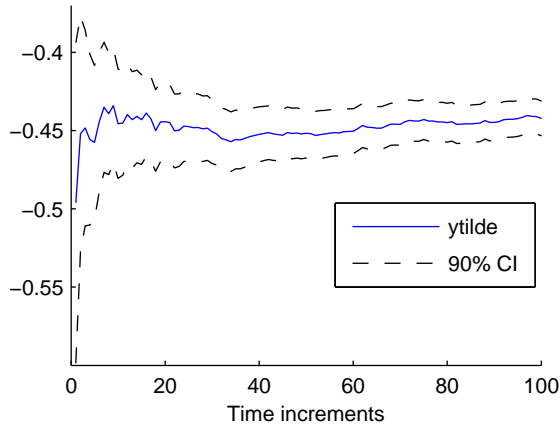


Figure 1: Convergence of the output of the MORET simulator for its nominal design values.

#### 2.4 The Kriging metamodel

Kriging is a functional approximation method originally coming from geosciences, and having been popularized in machine learning (Gaussian Process paradigm, see e.g. Rasmussen and Williams (2006)) and in numerous application fields. Kriging simultaneously provides an interpolator of the partially observed function  $y$ , the *Kriging mean predictor*  $m(\cdot)$ , and a measure of prediction uncertainty at every  $\mathbf{x}$ , the *Kriging variance*  $s^2(\cdot)$ . The basic idea is to see  $y$  as one realization of a square-integrable real-valued random process indexed by  $D$ , and to make optimal linear predictions of  $Y(\mathbf{x})$  given the  $Y$  values at the already evaluated input points  $\mathbf{X}^n := \{\mathbf{x}^i, 1 \leq i \leq n\}$ . Of course, this prediction depends on the two first moments of the process  $Y$ , which are generally assumed to be known up to some coefficients. Here we assume that  $Y$  has an unknown constant trend  $\mu \in \mathbb{R}$ , and a stationary covariance kernel  $k$ , i.e. of the form  $k : (\mathbf{x}, \mathbf{x}') \in D^2 \rightarrow k(\mathbf{x}, \mathbf{x}') = \sigma^2 r(\mathbf{x} - \mathbf{x}'; \psi)$  for some admissible correlation function  $r$  with parameters  $\psi$ . This is the framework of *Ordinary Kriging* (OK) (Matheron 1969). Additionally, assuming further that  $Y|\mu$  is a Gaussian Process (GP) and that  $\mu$  is independent of  $Y$  and follows an improper uniform distribution over  $\mathbb{R}$  leads to the convenient result that OK amounts to conditioning  $Y$  on the measurements, i.e. ensuring that  $m(\cdot)$  and  $s^2(\cdot)$  coincide respectively with the conditional mean and variance functions. We stick here to this set of assumptions, in order to get explicit (gaussian) conditional distributions for  $Y(\mathbf{x})$  knowing the observations, and to be in position to use generalizations of this to the heterogeneously noisy case.

Let us indeed come back to our noisy observations  $\tilde{y}_i = y(\mathbf{x}^i) + \epsilon_i$  ( $1 \leq i \leq n$ ). If we suppose that  $y$  is a realization of a GP following the OK assumptions above, the  $\tilde{y}_i$ 's can now be seen as realizations of the random variables  $\tilde{Y}_i := Y(\mathbf{x}^i) + \epsilon_i$ , so that Kriging amounts to conditioning  $Y$  on the heterogeneously noisy observations  $\tilde{Y}_i$  ( $1 \leq i \leq n$ ). As shown earlier in Ginsbourger, Picheny, Roustant and Richet (2008), provided that the process  $Y$  and the gaussian measurement errors  $\epsilon_i$  are stochastically independent, the process  $Y$  is still gaussian conditionally on the noisy observations  $\tilde{Y}_i$  ( $1 \leq i \leq n$ ), and its conditional mean and variance

functions are given by the following slightly modified OK equations:

$$m_n(\mathbf{x}) = \mathbb{E}[Y(\mathbf{x})|\widetilde{A}_n] = \widehat{\mu}_n + \mathbf{k}_n(\mathbf{x})^T (K_n + \Delta_n)^{-1} (\widetilde{\mathbf{y}}^n - \widehat{\mu}_n \mathbf{1}_n), \quad (3)$$

$$s_n^2(\mathbf{x}) = \text{Var}[Y(\mathbf{x})|\widetilde{A}_n] = \sigma^2 - \mathbf{k}_n(\mathbf{x})^T (K_n + \Delta_n)^{-1} \mathbf{k}_n(\mathbf{x}) + \frac{(1 - \mathbf{1}_n^T (K_n + \Delta_n)^{-1} \mathbf{k}_n(\mathbf{x}))^2}{\mathbf{1}_n^T (K_n + \Delta_n)^{-1} \mathbf{1}_n}, \quad (4)$$

where  $|\cdot|$  means "conditional on",  $\widetilde{\mathbf{y}}^n = (\widetilde{y}_1, \dots, \widetilde{y}_n)^T$ ,  $\widetilde{A}_n$  is the event  $\{Y(\mathbf{x}^i) + \varepsilon_i = \widetilde{y}_i, 1 \leq i \leq n\}$ ,  $K_n = (k(\mathbf{x}^i, \mathbf{x}^j))_{1 \leq i, j \leq n}$ ,  $\mathbf{k}_n(\mathbf{x}) = (k(\mathbf{x}, \mathbf{x}^1), \dots, k(\mathbf{x}, \mathbf{x}^n))^T$ ,  $\Delta_n$  is a diagonal matrix of diagonal terms  $\tau_1^2 \dots \tau_n^2$ ,  $\mathbf{1}_n$  is a  $n \times 1$  vector of ones, and  $\widehat{\mu}_n = \mathbf{1}_n^T (K_n + \Delta_n)^{-1} \widetilde{\mathbf{y}}^n / \mathbf{1}_n^T (K_n + \Delta_n)^{-1} \mathbf{1}_n$  is the best linear unbiased estimate of  $\mu$ .  $m(\cdot)$  and  $s^2(\cdot)$  are indexed by  $n$  in order to bring to light the dependence on the design of experiments, and to prepare the ground for the algorithmic developments needing sequential Kriging updates.

The only difference compared to OK equations is the replacement of  $K_n$  by  $K_n + \Delta_n$  at every occurrence. Specific properties of this generalization of OK include that  $m_n(\cdot)$  is not interpolating noisy measurements, that  $s_n^2(\cdot)$  doesn't vanish at that points and is globally inflated compared to the noiseless case. Note that although  $s_n^2(\cdot)$  now depends on both the design  $\mathbf{X}^n$  and the noise variances  $\boldsymbol{\tau}^2 := \{\tau_1^2, \dots, \tau_n^2\}$ , it still does not depend on the observations. Figure 2 shows an example of kriging based on noisy observations with heterogeneous noise. For small observation noise, the model is almost interpolant (e.g. at  $x = 0.5$ ), while for large noise the confidence interval remains large and the best predictor can be far from the observation.

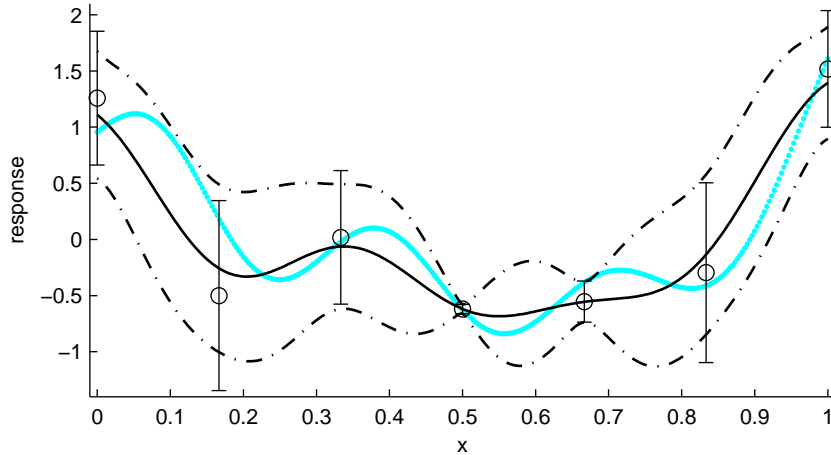


Figure 2: Actual function (bold gray), Kriging mean (bold black) and 90% confidence intervals (mixed line); the circles are the observation values  $\widetilde{y}_i$ , the bars show the noise amplitude ( $\pm 2 \times \tau_i$ ).

### 3. KRIGING-BASED OPTIMIZATION; LIMITATIONS WITH NOISY FUNCTIONS

Optimization (say minimization) based on Kriging with noiseless observations has truly become a hit following the publication of the EGO algorithm (Jones et al. 1998). EGO consists in sequentially evaluating  $y$  at a point maximizing a figure of merit relying on Kriging, the *Expected Improvement* criterion (EI), and updating the metamodel at each new observation. As illustrated in Jones (2001), directly minimizing  $m_n(\cdot)$  is inefficient since it may lead the sequence of good points to get trapped in an artificial basin of minimum,



whereas maximizing EI provides a right trade-off between exploitation and exploration in order to converge to a global minimizer. Our goal here is to adapt EI to the heterogeneously noisy case. Let us previously recall the definition and analytical expression of EI in the noiseless case.

Let  $y_i = y(\mathbf{x}^i)$  ( $1 \leq i \leq n$ ),  $\mathbf{y}^n = (y_1, \dots, y_n)^T$ ,  $A_n$  denote the event  $\{Y(\mathbf{x}^i) = y_i, 1 \leq i \leq n\}$ , and  $m_n$  and  $s_n^2$  still refer to the Kriging mean and variance. The idea underlying EI is that sampling at  $\mathbf{x}$  will bring an improvement of  $\min(y(\mathbf{X}^n)) - y(\mathbf{x})$  if  $y(\mathbf{x})$  is below the current minimum  $\min(y(\mathbf{X}^n))$ , and 0 otherwise. Of course, this quantity cannot be known in advance since  $y(\mathbf{x})$  is unknown. However, the GP model and the available information  $A_n$  make it possible to define and derive the following conditional expectation:

$$EI_n(\mathbf{x}) := \mathbb{E} \left[ (\min(Y(\mathbf{X}^n)) - Y(\mathbf{x}))^+ | A_n \right] = \mathbb{E} \left[ (\min(\mathbf{y}^n) - Y(\mathbf{x}))^+ | A_n \right] \quad (5)$$

An integration by parts yields the well-known analytical expression:

$$EI_n(\mathbf{x}) := (\min(\mathbf{y}^n) - m_n(\mathbf{x})) \Phi \left( \frac{\min(\mathbf{y}^n) - m_n(\mathbf{x})}{s_n(\mathbf{x})} \right) + s_n(\mathbf{x}) \phi \left( \frac{\min(\mathbf{y}^n) - m_n(\mathbf{x})}{s_n(\mathbf{x})} \right), \quad (6)$$

where  $\Phi$  and  $\phi$  are respectively the cumulative distribution function and the probability density function of the standard gaussian law. The latter analytical expression is very convenient since it allows fast evaluations of EI, and even analytical calculation of its gradient and higher order derivatives. This is used in particular in the DiceOptim R package (Roustant, Ginsbourger and Deville 2009) for speeding up EI maximization.

Let us now state why the classical EI is not well adapted to Kriging with noisy observations. Coming back to the previous notations, we have indeed:

$$EI_n(\mathbf{x}) = \mathbb{E} \left[ \left( \underbrace{\min(Y(\mathbf{X}^n))}_{\text{unknown}} - \underbrace{Y(\mathbf{x})}_{\text{unreachable}} \right)^+ \middle| \widetilde{A}_n \right], \quad (7)$$

which is not very satisfactory for at least two reasons. The first one is that the current minimum  $\min(Y(\mathbf{X}^n))$  is not deterministically known conditionally on the noisy observations, contrarily to the noiseless case. The second reason is that the EI is based on the improvement that could bring a deterministic evaluation of  $y$  at the candidate point  $\mathbf{x}$ . Now, if the next evaluation is noisy,  $Y(\mathbf{x})$  will remain non-exactly known. It would hence be more adapted to have a criterion taking the precision of the next measurement into account.

The first alternative to use the *EI* in the noisy case is to plug in a surrogate value for the unknown  $\min(Y(\mathbf{X}^n))$ . However, the natural choice  $\min(\widetilde{\mathbf{y}}^n)$  can be highly risky since the noisy minimum is a biased estimator of the noiseless minimum, and it suffices to have one highly noisy observation with a low value to deeply underestimate  $\min(\mathbf{y}^n)$  for the rest of the optimization.

A rule of thumb proposed by Vazquez, Villemonteix, Sidorkiewicz and Walter (2008) is to plug in the minimum of the Kriging mean predictor  $\min(m_n(\mathbf{X}^n))$  instead of  $\min(\widetilde{\mathbf{y}}^n)$ , which seems to be a more sensible

option in order to smooth out the noise fluctuations. In the same fashion, Huang, Allen, Notz and Zeng (2006) plug in the mean predictor at the training point with smallest kriging quantile.

However, using a plugin for the current minimum does not take into account the noise in the future response. In Huang, Allen, Notz and Zeng (2006), a variant of the EI called *Augmented Expected Improvement* (AEI) is proposed for uniformly noisy observations, the EI being multiplied by a penalization function to account for the diminishing return of observation replicates, which then takes into account the future noise, but at already sampled points only.

Forrester, Keane and Bressloff (2006) proposed a *reinterpolation technique* that replaces the noisy observations by the kriging mean predictor  $m_n(\mathbf{X}^n)$ , and fits a noise-free kriging on such data, which is used for the standard EGO algorithm. However, this heuristic was designed for deterministic errors and does not allow repeated observations, which can be desirable in our framework.

A more rigorous alternative consists of estimating the EI based on Monte-Carlo simulations involving the joint distribution of  $(\min(Y(\mathbf{X}^n), Y(\mathbf{x})))$  conditional on  $\widetilde{A}_n$ ; however, such estimates are noisy and numerically costly, which makes the EI maximization challenging.

In the next section, we present an alternative infill criterion that takes into account past and future noises with transparent probabilistic foundations, and which can be derived analytically.

#### 4. QUANTILE-BASED EXPECTED IMPROVEMENT

We now introduce a variant of EI for the case of a deterministic objective function with noisy measurements with heterogeneous variances. Our aim is to get a Kriging-based optimization criterion measuring which level of improvement can be statistically expected from sampling  $y$  at a new  $\mathbf{x}$  with a noise of given variance  $\tau^2$ . A first question to be addressed is of decision-theoretic nature: what does the term "improvement" mean when comparing two sets of noisy observations? According to what kind of criterion should we judge that a set of noisy observations, or the associated metamodel, is better (in terms of minimization) after the  $(n + 1)^{\text{th}}$  measurement than before it?

Obviously, using only the noisy observations  $\widetilde{\mathbf{y}}^n$  and  $\widetilde{y}^{n+1}$  for that matter is a highly risky strategy, since the noise may introduce errors in the ranking of the observations. Here we propose to use the  $\beta$ -quantiles given by the Kriging conditional distribution, for a given level  $\beta \in [0.5, 1[$ : a point is declared "best" over a set of candidates  $\mathbf{X}^n$  whenever it has the lowest  $\beta$ -quantile:

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x} \in \mathbf{X}_n} q_n(\mathbf{x}) \tag{8}$$

with  $q_n(\mathbf{x}) := \inf\{u \in \mathbb{R} : \mathbb{P}(Y(\mathbf{x}) \leq u | \widetilde{A}_n) \geq \beta\} = m_n(\mathbf{x}) + \Phi^{-1}(\beta)s_n(\mathbf{x})$ , which is the decision criterion also used in Huang, Allen, Notz and Zeng (2006).

Now, we propose to define an improvement that is *consistent* with our decision criterion: improvement

$I$  is here the decrease of the lowest  $\beta$ -quantile, between the present step  $n$  and the forthcoming step  $n + 1$ :

$$I = \left( \min(q_n(\mathbf{X}_n)) - q_{n+1}(\mathbf{x}^{n+1}) \right)^+ \quad (9)$$

Of course, like in the noiseless case, this improvement cannot be known in advance, because  $q_{n+1}(\mathbf{x}_{n+1})$  depends on the future observation  $\tilde{y}_{n+1}$ . However, thanks to the particular form of the kriging equations, the future quantile  $q_{n+1}$  can be predicted, and consequently the EI calculated, based on the GP model at step  $n$ , as we show below.

One can remark that we restrict here our choice to the observed points ( $\mathbf{X}_n$  and  $\mathbf{x}^{n+1}$ ), even though a similar criterion could be defined over the entire design space:  $I = (\min_D(q_n(\mathbf{x})) - \min_D(q_{n+1}(\mathbf{x})))^+$ . Such a restriction greatly simplifies calculations, and is a reasonable conservative measure: indeed, who would trust a metamodel so much to propose a final candidate minimizer without any measurement at that point?

Let us denote by  $Q_i(\mathbf{x})$  the kriging quantile  $q_i(\mathbf{x})$  ( $i \leq n + 1$ ) where the measurements are still in their random form, and define the Expected Quantile Improvement (EQI) as:

$$EQI_n(\mathbf{x}^{n+1}, \tau_{n+1}^2) := \mathbb{E} \left[ \left( \min_{i \leq n} (Q_n(\mathbf{x}^i)) - Q_{n+1}(\mathbf{x}^{n+1}) \right)^+ \middle| \widetilde{A}_n \right] \quad (10)$$

where the dependence on the future noise  $\tau_{n+1}^2$  appears through  $Q_{n+1}(\mathbf{x})$ 's distribution. The randomness of  $Q_{n+1}(\mathbf{x})$  conditional on  $\widetilde{A}_n$  is indeed a consequence from  $\widetilde{Y}_{n+1} := Y(\mathbf{x}^{n+1}) + \varepsilon_{n+1}$  having not been observed yet at step  $n$ . However, following the fact that  $\widetilde{Y}_{n+1} | \widetilde{A}_n$  is gaussian with known mean and variance, one can show that  $Q_{n+1}(\cdot)$  is a GP conditional on  $\widetilde{A}_n$  (see proof and details in appendix). Furthermore,  $\min_{i \leq n} (Q_n(\mathbf{x}^i))$  is known conditional on  $\widetilde{A}_n$ . As a result, the proposed  $EQI$  is analytically tractable, and we get by a similar calculation as in Eq. 6:

$$EQI_n(\mathbf{x}^{n+1}, \tau_{n+1}^2) = (\min(\mathbf{q}^n) - m_{Q_{n+1}}) \Phi \left( \frac{\min(\mathbf{q}^n) - m_{Q_{n+1}}}{s_{Q_{n+1}}} \right) + s_{Q_{n+1}} \phi \left( \frac{\min(\mathbf{q}^n) - m_{Q_{n+1}}}{s_{Q_{n+1}}} \right) \quad (11)$$

where:  $\left\{ \begin{array}{l} \mathbf{q}^n := \{q_n(\mathbf{x}^i), i \leq n\}$  is the set of current quantile values at the already visited points,  
 $m_{Q_{n+1}} := \mathbb{E}[Q_{n+1}(\mathbf{x}^{n+1}) | \widetilde{A}_n]$  is  $Q_{n+1}(\mathbf{x}^{n+1})$ 's conditional expectation —seen from step  $n$ ,  
 $s_{Q_{n+1}}^2 := \text{Var}[Q_{n+1}(\mathbf{x}^{n+1}) | \widetilde{A}_n]$  is its conditional variance, both derived in appendix.

As in the noiseless case, the EQI criterion is hence known in closed form, which is a desirable feature for its maximization.  $\tau^2$  and  $\beta$  are to be considered here as parameters, and EQI maximization is done with respect to  $\mathbf{x}^{n+1}$  only.

The EQI criterion has the following important properties:

- in absence of future noise ( $\tau_{n+1}^2 = 0$ ), the future quantile at  $\mathbf{x}^{n+1}$  coincides with the observation

$\tilde{y}^{n+1} = y(\mathbf{x}^{n+1})$ , since the prediction variance at this point will be null; it follows directly that  $Q_{n+1}(\mathbf{x}^{n+1})|\widetilde{A}_n = Y_{n+1}|\widetilde{A}_n$ , so the EQI is then equal to the classical EI with a plugin of the kriging quantiles for  $\min(\mathbf{y}^n)$

- in absence of past noise (for the  $n$  first observations),  $\min(\mathbf{q}^n)$  is equal to the minimum of the observations,  $\min(\mathbf{y}^n)$
- in absence of both past and future noise, the EQI is then equal to the classical EI.

$\beta$  tunes the level of reliability wanted on the final result; with  $\beta = 0.5$ , the design points are compared based on the kriging mean predictor only without taking into account the prediction variance at those points, while high values of  $\beta$  (i.e. near to 1) penalize designs with high uncertainty, which is a more conservative approach. Hence, with a high  $\beta$ , the criterion is more likely to favour observation repetitions or clustering, in order to locally decrease the prediction variance, while with  $\beta = 0.5$ , the criterion can be expected to be tendentially more exploratory.

The future noise  $\tau_{n+1}^2$  also strongly affects the shape of the EQI. Indeed, a very noisy future observation can only have a very limited influence on the kriging model, so the only possibility to have a non-null improvement is either to sample where  $q_n(\mathbf{x})$  is minimum if this point has not been observed yet, which will bring an improvement even if  $q_{n+1}(\mathbf{x}^{n+1}) = q_n(\mathbf{x}^{n+1})$ , or to sample at the current best point, which may decrease its uncertainty and brings a small but measurable improvement. On the contrary, if  $\tau_{n+1}^2$  is very small, the EQI behaves like the classical EI, making the well-known trade-off between exploration and exploitation.

Figure ?? illustrate the dependence of the EQI on both  $\beta$  and  $\tau_{n+1}^2$ .

The proposed EQI criterion seems to be quite conservative compared to other EI variants, since it intends to measure the *actual* improvements on the final decision criterion, which can be very limited for a single step. In the next section, we propose a numerical trick that turns this property into a substantial asset, and results in an optimization strategy less myopic than the classical EGO.

## 5. OPTIMIZATION WITH FINITE COMPUTATIONAL BUDGET

It is well-known that the EGO algorithm is a so-called *myopic* strategy, since its criterion EI always considers the next step as if it were the last one. However, for most computer experiments, the total computational budget is bounded, and prescribed by industrial constraints such as time and power limitations. In the deterministic framework, this results in a limited (given) number of observations for optimization. It has been shown in that case in Mockus (1988) followed by Ginsbourger and Le Riche (2009) that taking into account the finite budget may modify the optimization strategy and improve significantly its efficiency.

In our framework of simulators with tunable precision, the concept of finite budget is particularly critical, since each observation requires a trade-off between accuracy and rapidity, and in general, the user has to

trade off between the total number of observations and their precision. This problem is typical of the theory of optimal designs in linear modeling (Fedorov and Hackl 1997), with the notable difference that we face it here within a sequential strategy.

Coming back to the notations of section 2, we defined a noisy measurement as depending on both a location  $\mathbf{x}_i$  and a computational time  $t_i$ , and which variance is given by  $\tau^2(t_i)$  (for the clarity of the writing the potential dependence on  $\mathbf{x}_i$  is not considered here). At step  $n$ , the remaining budget for optimization is  $T_{n+1} = T_0 - \sum_{i=1}^n t_i$ . Then, the computational constraint implies that the sum of all computational times is fixed to a given budget, say  $T_0$ . In the case of Monte Carlo-based simulators, this computational budget can be defined in terms of sample size. i.e. total number of drawings.

The EQI criterion allows taking into account such computational boundedness in the choice of the new candidate observations. Indeed, the future noise level  $\tau_{n+1}^2$ , which is a parameter of the EQI criterion, will stand here for the finite resource. Given a computational budget  $T_{n+1}$ , the smallest noise variance achievable for a new measurement is  $\tau^2(T_{n+1})$ , assuming that all the remaining budget will be attributed to this measurement. In the course of the optimization process, the remaining budget decreases, so  $\tau^2(T_{n+1})$  increases with  $n$ .

Then, we propose to set  $\tau_{n+1}^2 = \tau^2(T_{n+1})$  for the EQI calculation within our sequential optimization procedure, meaning that the EQI will measure the potential improvement if all the remaining budget is attributed to the next observation. Of course, the actual budget for the next observation may be a lot smaller than  $T_{n+1}$  so the optimization does not stop after one step. With this setting, the new experiment is chosen knowing that even if all the budget is used for a single observation, its noise variance will not decrease below a certain value.

Consequently, the EQI will behave differently at the beginning and at the end of the optimization. When the budget is high, EQI will be highest in unexplored regions, since it is where accurate measurements are likely to be most efficient (the EQI will actually be almost similar to a classical EI). At the end of the optimization, however, when the remaining time is small, the EQI will be small in unexplored regions since even if the actual function is low, there is not enough computational time to obtain a lower quantile than the current best one. In that case, the EQI will be highest close to the current best point(s).

## 6. ALLOCATION OF RESOURCE

In the previous sections, we proposed an infill criterion for choosing sequentially sample points, and a specific tuning of the criterion to account for finite computational resources. However, the question of the choice of the computational budget for a single observation has yet been left open. This section proposes two answers, the first by using the simple strategy of fixed budget, the second by taking advantage of the response convergence monitoring to dynamically adapt the budget to each measurement.

## 6.1 Constant allocation

We assume first that the computational budget  $T_0$  can be divided in elementary time steps  $t_e \in ]0, +\infty[$ , so that  $T_0 = N_0 \times t_e$ . An elementary step can correspond for example to a given number of solver iterations for partial convergence, or a number of drawings for stochastic simulators. The algorithm with constant allocation will then attribute one by one the  $N_0$  elementary time steps to either generate new measurements or improve accuracy on existing ones.

At step  $n$ , a budget  $n \times t_e < T_0$  as already been spent on the measurements. The new measurement is to be chosen using the EQI criterion. At unsampled locations, the criterion is simply evaluated with  $\tau_{n+1}^2 = \tau^2(T_{n+1})$  (where  $T_{n+1} = T_0 - nt_e$ ). At existing observations, a different value has to be used; if not, the EQI would estimate the value of a new measurement with variance  $\tau^2(T_{n+1})$ , instead of the value of improving the existing measurement.

To compute this value, we use the fact that it is equivalent for the kriging model to have at the same point several measurements with independent noises or a single equivalent measurement that is the weighted average of the observations. For instance, let  $\tilde{y}_{i,1}$  and  $\tilde{y}_{i,2}$  be two measurements with respective noise levels  $\tau_{i,1}^2$  and  $\tau_{i,2}^2$ . They are equivalent to a single measurement

$$\tilde{y}_{i,eq} = \frac{\tau_{i,1}^{-2} \tilde{y}_{i,1} + \tau_{i,2}^{-2} \tilde{y}_{i,2}}{\tau_{i,1}^{-2} + \tau_{i,2}^{-2}} \quad (12)$$

with variance  $\tau_{i,eq}^2$  the harmonic mean of  $\tau_{i,1}^2$  and  $\tau_{i,2}^2$ , namely:

$$\frac{1}{\tau_{i,eq}^2} := \frac{1}{\tau_{i,1}^2} + \frac{1}{\tau_{i,2}^2} \implies \tau_{i,eq}^2 = \frac{\tau_{i,1}^2 \tau_{i,2}^2}{\tau_{i,1}^2 + \tau_{i,2}^2} \quad (13)$$

Now, for the EQI calculation, we make the assumption that it is equivalent to carry the measurement process from zero until the noise level  $\tau_i^2[b_i]$  is reached, or to make a new measurement with noise variance:

$$\tau_i^2[j \rightarrow b_i] := \frac{\tau_i^2[j] \tau_i^2[b_i]}{\tau_i^2[j] - \tau_i^2[b_i]} = \frac{\tau^2(j \times t_e) \tau^2(T_i)}{\tau^2(j \times t_e) - \tau^2(T_i)} =: \tau^2(j \times t_e \rightarrow T_i) \quad (14)$$

If  $j = 0$ , we have  $\tau^2(j \times t_e \rightarrow T_i) = \tau^2(T_i)$ .

Once the new point  $\mathbf{x}^{n+1}$  is chosen and the measurement is made, the kriging model has to be updated. If  $\mathbf{x}^{n+1} \notin \mathbf{X}_n$ , the new point has to be added to the DoE and the kriging equations modified accordingly; otherwise, it only requires to replace the previous values of response and noise  $\tilde{y}_i[j-1]$  and  $\tau_i^2[j-1]$  by  $\tilde{y}_i[j]$  and  $\tau_i^2[j]$  in the kriging equations 4. The remaining time is then updated and the next point can be chosen. The algorithm with constant allocation is presented in pseudo-code form in table 1.

Table 1: EQI algorithm with constant allocation

---

- Choose  $T_0, \beta, t_e, n_0$
- Build initial DoE  $\mathbf{X}^{n_0}$ , generate observations  $\tilde{\mathbf{y}}^{n_0}$ , fit Kriging model
- Set  $n = n_0$
- while**  $n < N$ 
  - Choose new design point  $\mathbf{x}^{n+1}$  that maximizes  $EQI_n(\cdot, \tau_{n+1}^2)$ , with:
    - $\tau_{n+1}^2 = \tau^2(t_{n+1})$  if  $\mathbf{x} \notin \mathbf{X}_n$
    - $\tau_{n+1}^2 = \tau^2(b_i \times t_e \rightarrow t_{n+1})$  if  $\mathbf{x} = \mathbf{x}^i$
  - If**  $\mathbf{x}_{n+1} \in \mathbf{X}_n$ :
    - Update corresponding  $\tilde{y}_i$  and  $\tau_i^2$
  - Otherwise**
    - Generate  $\tilde{y}_{n+1}$ , set  $\tau_{n+1}^2 = \tau^2(t_e)$
    - Set  $n = n + 1$
  - Update kriging model
  - **end if**
- end while**
- Choose  $\mathbf{x}^*$  based on the Kriging quantiles at the measurement points.

---

## 6.2 Repeatable experiments with homogenous noise

The popular framework of repeatable experiments with homogenous noise can be seen as a particular case of constant allocation, and is worth expliciting since it allows great simplifications of the above equations. We consider that each observation requires a time  $t_e$ , and has a noise variance  $\tau^2$ .

Although the variance is constant, we need to define a variance depending on computational time to account for the finite budget. Here, attributing all the remaining budget to a new measurement at  $\mathbf{x}^{n+1}$  means performing all the remaining  $N_0 - n$  measurements at this point. As mentioned earlier, it is equivalent, in terms of kriging, to have  $k$  measurements  $\tilde{y}_{i,1} \dots \tilde{y}_{i,k}$  at  $\mathbf{x}^i$  or a single measurement  $\tilde{y}_{i,eq} = \frac{1}{k} \sum_{j=1}^k (\tilde{y}_{i,j})$  with variance  $\tau_{i,eq}^2 = \tau^2/k$ . Improving a measurement is simply updating the equivalent observation with a new repetition, i.e.  $\tilde{y}_{i,eq} \leftarrow \frac{1}{k+1} (k \times \tilde{y}_{i,eq} + \tilde{y}_{i,k+1})$ , or equivalently, add this observation to the DOE. Then,  $EQI$  writes similarly whether it is evaluated at an existing observation or not, and the future noise is set to  $\tau^2/(N_0 - n)$ .

The algorithm for repeatable experiments is presented in pseudo-code form in table 2. In this version, possibly repeated experiments are not aggregated into an equivalent observation, which makes the implementation simpler, but might increase the cost for the kriging evaluations since the covariance matrix will be larger.

Table 2: EQI algorithm with repeatable experiments with homogeneous noise

---

- Choose  $T_0, \beta, t_e, n_0$
- Build initial DoE  $\mathbf{X}^{n_0}$ , generate observations  $\tilde{\mathbf{y}}^{n_0}$ , fit Kriging model
- for**  $n$  **from**  $n_0$  **to**  $N_0$ 
  - Choose new design point  $\mathbf{x}^{n+1}$  that maximizes  $EQI_n(\cdot, \tau^2/(N_0 - n))$
  - Generate  $\tilde{y}^{n+1}$
  - Update kriging model with  $\tilde{y}^{n+1}, \mathbf{x}^{n+1}$  and  $\tau^2$
- end while**
- Choose final design based on the Kriging quantile

---

### 6.3 On-line allocation

The constant allocation strategy of the previous section performs  $N - N_0$  EGO iterations, and each requires running an inner optimization loop for the maximization of the EQI, which can be very time-consuming. Hence, the elementary time step  $t_e$  must be chosen large enough to limit the number of EQI optimizations. Typically, with partial convergence or stochastic simulators,  $t_e$  must be chosen a lot larger than a single solver iteration or drawing, respectively. This limitation can greatly hinder the flexibility and potential of tunable precision, since it reduces the possibilities of a quasi-continuum of fidelities to a few discrete precision levels.

In this section, we propose to overcome this problem by proposing a heuristic for dynamically choosing the computational resource given to an experiment. A simple way to do so is to monitor the evolution of the EQI at the current observation point. Indeed, instead of maximizing the EQI after each  $t_e$  is spent, we will choose a observation point, and allocate several time steps on it until a criterion is met. As for the constant allocation case, the EQI is updated after each step, by replacing the previous values of response and noise  $\tilde{y}_i[j-1]$  and  $\tau_i^2[j-1]$  by  $\tilde{y}_i[j]$  and  $\tau_i^2[j]$  in the kriging equations, and by replacing the future noise level  $\tau^2(T_i)$  by  $\tau^2(T_i - t_e)$ .

The updated EQI tends by construction to decrease when computation time is added, since first the kriging uncertainty reduces at the observation point and second since EQI decreases when  $\tau^2(j \times t_e \rightarrow T_i)$  increases. However, if the measurement converges to a good (small) value, EQI can increase temporarily. Inversely, if the measurement converges to a high value, EQI decreases faster. Hence, we can define a ("point switching") stopping criterion for resource allocation based on EQI. If the EQI decreases below a certain value, carrying on the calculations is not likely to help the optimization, so the observation process should stop and another point be chosen. Here, we propose to interrupt a measurement and search for a new point when the current value of the EQI is less than a proportion of the initial EQI value (that is, the value of EQI when starting the measurement process at that point), for instance 50%.

The sequence of this new procedure is as follow: first, choose the point with highest expected quantile improvement given the whole remaining budget, store the corresponding EQI value as reference, and then invest new elementary measurements at this point until the EQI with updated data falls under a given proportion  $\gamma \in ]0, 1[$  of the reference EQI value. The operation of choosing the most promising point is then started again, and so on until the total computational budget has been spent. Note that the final number of measurements and EQI maximizations are not determined beforehand but adapts automatically to the budget and resource distribution. The algorithm is presented in pseudo-code form in table 3.

### 6.4 Practical issues

The total computational budget  $T$  needs to be defined before optimization, and discretized in incremental steps  $\{t_e, 2 \times t_e, \dots, b \times t_e\}$ , where  $b = T/t_e$ . Smaller steps (i.e. a smaller  $t_e$ ) result in increased precision,



Table 3: Quantile EI algorithm with on-line resource allocation

---

- Build initial DoE  $\mathbf{X}^{n_0}$ , generate observations  $\tilde{\mathbf{y}}^{n_0}$  using  $T_0$  computational time, fit Kriging model
- Set  $n = n_0$  and  $T_n = T - T_0$

**while**  $T_n > 0$

- Choose new design point  $\mathbf{x}^{n+1}$  that maximizes  $EQI_n(\cdot, \tau^2(T_n))$
- Generate  $\tilde{y}_{n+1}[1]$  with one time increment
- Augment DoE:  $\mathbf{X}^{n+1} = \{\mathbf{X}^n, \mathbf{x}^{n+1}\}$
- Update Kriging model with  $\tilde{y}_{n+1} = \tilde{y}_{n+1}[1]$  and  $\tau_{n+1}^2 = \tau^2(t_e)$
- Set  $T_{n+1} = T_n - t_e$ ,  $j = 1$ , and  $t = t_e$

**while**  $EQI_{n+1}(\mathbf{x}^{n+1}, \tau^2(t \rightarrow T_n)) > \gamma EQI_n(\mathbf{x}^{n+1}, \tau^2(T_n))$

- Generate  $\tilde{y}_{n+1}[j+1]$  by adding one time increment
- Set  $T_{n+1} = T_{n+1} - t_e$ ,  $j = j + 1$ , and  $t = t + t_e$
- Update Kriging model with:  $\tilde{y}_{n+1} = \tilde{y}_{n+1}[j]$ ,  $\tau_{n+1}^2 = \tau^2(t)$

**end while**

- Set  $n = n + 1$

**end while**

- Choose final design based on the Kriging quantile

---

but requires more Kriging updates and EQI maximization, which can become computationally intensive. A prescribed fraction  $T_0$  of this budget is allocated to build an initial DoE, which should be designed in order to fit a realistic Kriging model. Based on previous numerical experiments, it has been found that using 10% to 30% of the total budget on a space-filling DoE (for instance, an LHS design) with uniform observation variances is a reasonable option.

Although the EQI criterion is analytical, its maximization with respect to  $\mathbf{x}^{n+1}$  is potentially complex and time-consuming. Indeed, the EQI (like the classical EI) is highly multimodal and require the use of global search algorithms (population-based techniques for instance). Also, each EQI evaluation requires the inversion of a  $(n + 1) \times (n + 1)$  covariance matrix, which can be expensive for large  $n$ . A valuable computational shortcut can be achieved in the update of the inverse of the covariance matrix when adding an observation using Schur's complement formula (see Marrel (2008)). It is to be noted that the gradients of EQI can be calculated analytically for a given covariance kernel (in the fashion of Ginsbourger (2009) (Chapter 4)).

On the three algorithms proposed in this section, the variance level depends only on computational time; however, the algorithms write similarly if the variance is design-dependent, by replacing  $\tau^2(t)$  by  $\tau^2(t; \mathbf{x})$ . Although not considered in the rest of this article, it is the authors' belief that such framework would be particularly adapted for the EQI criterion, since it would automatically take the additional uncertainty of certain measurements into account for the optimization strategy.

Also, a minimum achievable noise can be set by the user or the simulator itself. In that case,  $\tau^2$  should be bounded by a  $\tau_{min}^2 = \tau^2(T_{max})$ , and  $T_n$  replaced by  $\min(T_n, T_{max})$  in all  $EQI$  expressions. In that case, we would have  $\tau^2(t \rightarrow T_i) = \tau^2(t \rightarrow T_{max})$  for  $T_i > T_{max}$ , and  $\tau^2(T_{max} \rightarrow T_i) = +\infty$ , which would result in  $EQI = 0$ , so once  $T_{max}$  time would have been spent on an observation, it would never be chosen again.

Finally, a crucial point is that the EQI criterion requires that the relation between the error variance and computational time is known. As noted in section 2, a preliminary study may be needed to calibrate an error model, for instance by studying the convergence of a small number of simulations and inferring a parametric model for the noise. Those simulations can then be integrated in the initial DOE. Such error model might of course be updated during the optimization with the help of the new observations.

## 7. EXPERIMENTS

Two examples are proposed in this section: a one-dimensional analytical example, for illustrative purpose, and a two-dimensional nuclear safety problem.

### 7.1 One-dimensional example

First, we study a one-dimensional problem, with objective function defined over  $[0, 1]$  by:

$$y(x) = \frac{1}{2} \left( \frac{\sin(20x)}{1+x} + 3x^3 \cos(5x) + 10(x-0.5)^2 - 0.6 \right) \quad (15)$$

The noise is here inversely proportional to computational time, and independent of  $\mathbf{x}$ :  $\tau^2(t) = 0.1/t$ .

First, we represent the EQI criterion (with  $\beta = 0.9$ ) for the initial DoE and kriging model, for three different  $\tau^2$  values : 1, 0.1 and 0.01. The initial DoE consists of five equally-spaced measurements, with noise variances equal to 0.02 (the 95% confidence interval at a measurement point is approximately 25% of the range of  $y$ ). The kriging model has a gaussian covariance kernel with parameters  $\sigma = 1$  and  $\theta = 0.1$ . The true function, kriging model, and EQI are shown on Figure 3.

We can see that the choice of the future noise level has a great influence on the criterion. With small noise variance, the EQI behaves like the classical EI, with highest values in regions with high uncertainty and low mean predictions. With higher noise variances, the criterion becomes very conservative since is it high only in the vicinity of existing measurements. Indeed, a very noisy future observation can only have a very limited influence on the kriging model, so adding such observation in an uncertain region is insufficient to lower enough the quantile to have a high EQI.

Then, an optimization is performed with a total computational budget of  $T = 100$ , starting from the DoE described above.  $T$  is divided in 100 time increments. Each initial DoE measurement has required five time units ( $t_e = 1$ ), so the DoE used 25% of the computational budget. Here the kriging parameters are assumed to be known and are not re-evaluated at each iteration.

Figure 4 represents the final DoE and kriging model. Nine measurement points have been added, with computational times varying from one to 41. The final DoE consists of highly noisy observations space-filling the design region and a cluster of accurate observations in the region of the global optimum.

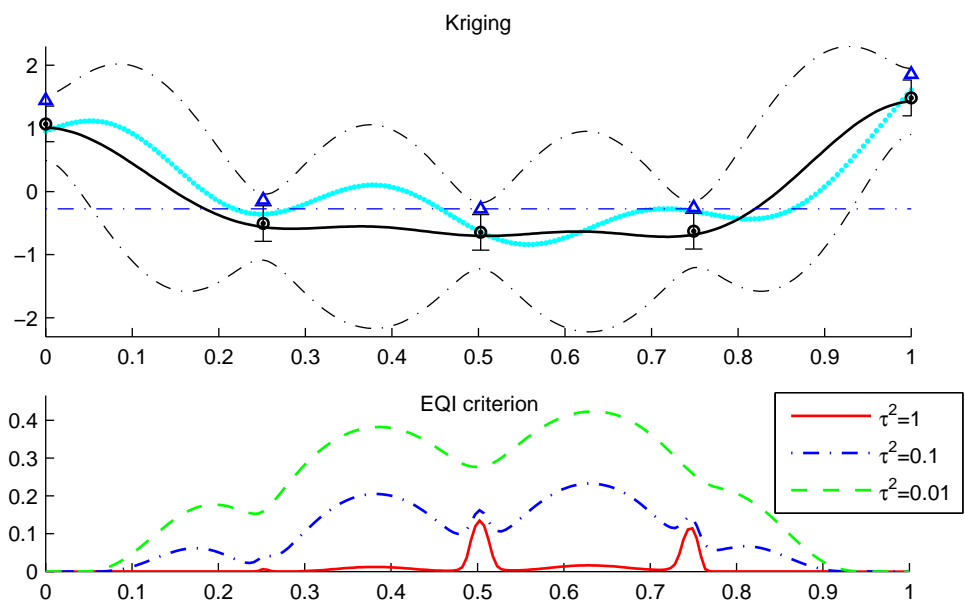


Figure 3: *Upper graph*: Actual function and Kriging. The horizontal line is the current best quantile. *Lower graph*: Corresponding EQI for three different future noise levels.

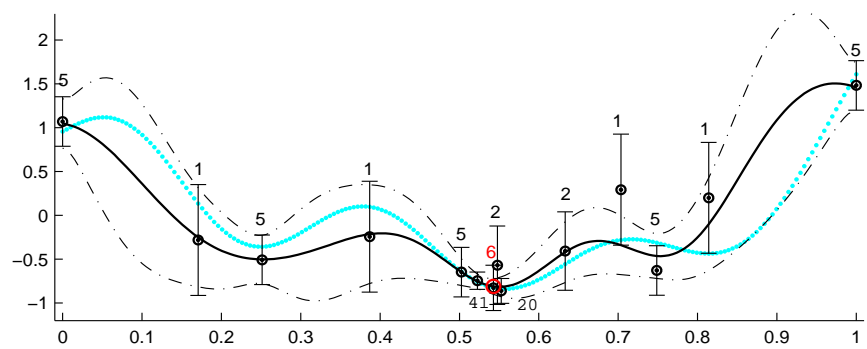


Figure 4: Observations and kriging after optimization. The numbers are the times steps for each observation. The chosen best observation is represented with the larger circle.

## 7.2 Application to a 2D benchmark from nuclear criticality safety assessments

In this section, the optimization algorithm is applied to the problem of safety assessment of a nuclear system involving fissile materials. The benchmark system used is an interim storage of dry  $PuO_2$  powder into a regular array of storage tubes. The criticality safety of this system is evaluated through the neutron multiplication factor (called k-effective or  $k_{\text{eff}}$ ), which models the nuclear chain reaction trend:

- $k_{\text{eff}} > 1$  is an increasing neutrons production leading to an uncontrolled chain reaction,
- $k_{\text{eff}} = 1$  means a stable neutrons population as required in nuclear reactors,
- $k_{\text{eff}} < 1$  is the safety state required for all unused fissile materials, like for fuel storage.

The neutron multiplication factor depends on many parameters such as the composition of fissile materials, operation conditions, geometry, etc. For a given set of parameters, the value of  $k_{\text{eff}}$  can be evaluated using the MORET stochastic simulator (Fernex et al. 2005), which is based on Markov Chain Monte-Carlo (MCMC) simulation techniques. The precision of the evaluation depends on the amount of simulated particles (neutrons), which is tunable by the user.

When assessing the safety of a system, one has to ensure that, given a set of admissible values  $D$  for the parameters  $\mathbf{x}$ , there are no physical conditions under which the  $k_{\text{eff}}$  can reach the critical value of 1.0 (minus a margin, usually chosen as 0.05):

$$\max_{\mathbf{x} \in D} k_{\text{eff}}(\mathbf{x}) \leq 1.0 - \text{margin} \quad (16)$$

The search for the worst combination of parameters  $\mathbf{x}$  defines a noisy optimization problem which is often challenging in practice, due to the possible high computational expense of the MORET simulator. An efficient resolution technique of this problem is particularly crucial since this optimization may be done numerous times. To account for the noise, classically the actual  $k_{\text{eff}}$  is replaced by its estimate  $\hat{k}_{\text{eff}}$  plus three output standard deviations. Here this conservative measure is replaced by the kriging quantile approach.

In this article, we focus on the maximization of  $k_{\text{eff}}$  with respect to two parameters, the other possible inputs being fixed to their most penalizing values (based on expert knowledge):

- $d.puo2$ , the density of the fissile powder, with original range  $[0.5, 4] \text{ g.cm}^{-3}$ , rescaled to the  $[0, 1]$  interval,

- $d.brouiscale$ , the density of water between storage tubes, with range  $[0, 1]$ , which accounts for the possible flooding of the storage (leading to an interstitial moderation of the neutrons interacting from a storage tube to another one).

Hence, to agree with previous notations, we set:  $\mathbf{x} = (d.puo2, d.brouiscale)$ , and  $y(\mathbf{x}) = -k_{\text{eff}}(\mathbf{x})$ . Simulation time is assumed proportional to the number of particles simulated (the entry cost of a new simulation being neglected). Since the simulator is based on MCMC method, the variance of the  $k_{\text{eff}}$  estimate is exactly inversely proportional to the number of particles. The variance slightly varies with input parameters, but this dependence can be considered negligible here.

For practical considerations, the optimization space  $D$  is discretized in a  $75 \times 75$  grid, and for each new measurement the EQI maximization is performed by exhaustive search on the grid. The incremental time step  $t_e$  is defined by the simulation of 4000 particles, which takes about half of minutes on a 3 GHz CPU. The response noise standard deviation can take values between  $5.23 \times 10^{-2}$  (one time step) and  $4.01 \times 10^{-3}$  (200 time steps).

To evaluate the efficiency of our algorithm, all the 5625 points of the grid have been evaluated with highest precision, which gives an accurate estimation of the shape (represented in figure 5), minimal value and minimizer  $x^*$  of the function. With this accurate dataset we find that  $x^* = [0.1892, 0.0811]$  and  $f(x^*) = -0.9847$ .

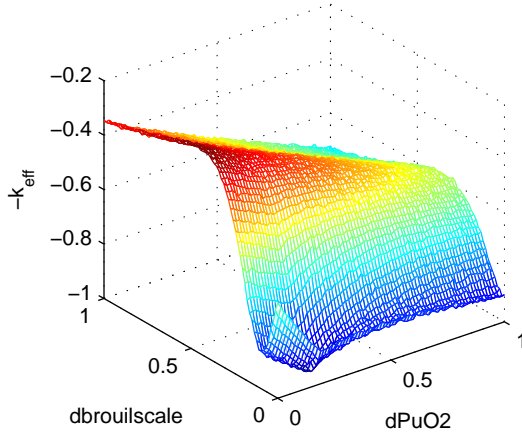


Figure 5: Accurate evaluation of  $k_{\text{eff}}$  (based on 200 time steps for each point) as a function of the design variables.

The  $k_{\text{eff}}$  range is approximately  $[0.3, 1.0]$ , so with one time step, the measurement 95% confidence interval length is  $4 \times 0.0523 = 0.209$ , which is about 30% of the response range. With 200 time steps, the length is 2% of the range. We consider two computational budgets here:  $T = 30$  and  $T = 100$ , which correspond to a single observation with standard deviation respectively equal to  $7.3 \times 10^{-3}$  and  $5.7 \times 10^{-3}$ . Both can be considered as very small budgets regarding the problem complexity. The initial DoEs consist of respectively 6 and 20-point random designs (with optimized *maximin* distance), with one time step used for each measurement (so respectively 17 and 20% of the budget is allocated to the initial DoE).

The kriging fit is made using the R package DiceKriging (Roustant et al. 2009). The chosen model has a constant trend (ordinary kriging) and Matern anisotropic covariance function. The covariance parameters are re-evaluated after each new observation.

The optimization results are given in tables 4 and 5. Figures 6 and 7 show the final kriging after optimization (mean, standard deviation and 90<sup>th</sup> quantile); figure 8 represents the final designs of experiments along with contour lines of the actual response.

For the budget  $T = 30$ , during optimization ten measurements have been added with time steps varying

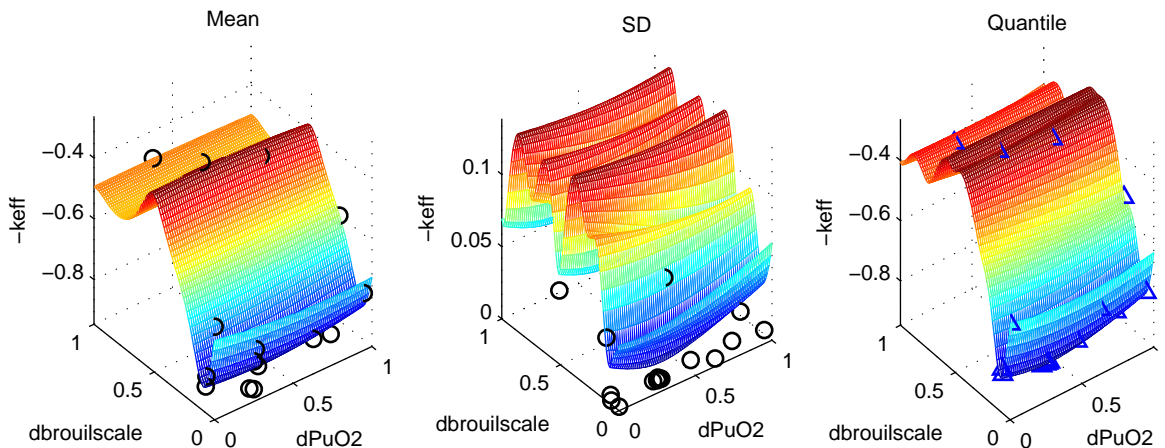


Figure 6: Final kriging after optimization for a  $T = 30$  budget.

from one to six. The best design  $\mathbf{x} = [0.3378, 0.0811]$  (based on the kriging quantile) has a kriging standard deviation of 0.0123 and is relatively close to the actual minimizer. From figure 8 (left), we see that five points were used for exploration (with only one time step) and the other five form a cluster of more accurate points in the optimal region. With such a low budget, the resulting kriging model (figure 6) is quite imprecise since the best predictor differs substantially in shape from the actual function and the kriging standard deviation is high over most of the design region. However, the points added during optimization reduce the error locally which results in a low standard deviation at the best design.

For the budget  $T = 100$ , during optimization 14 measurements have been added with time steps varying from one to 36. The best design  $\mathbf{x} = [0.1892, 0.0676]$  has a kriging standard deviation of 0.0071 and is almost equal to the actual minimizer. Here, approximately one third of the computational budget is allocated to the best design itself. The final kriging (figure 7) is more accurate than in the previous case, even though the standard deviation remains high in all the regions with high response values. In the region of the optimum, the kriging quantile is almost similar to the actual function.

Table 4: Sequential measurements for a  $T = 30$  budget. (bold: best value)

Iteration	$\mathbf{x}$	Time allocated	$\tilde{y}$	Kriging SD	Kriging quantile	Distance to actual optimum
0	[0.3108, 0.5135]	1	-0.3360	0.0515	-0.2920	0.4492
0	[0.6486, 0.0270]	1	-0.8462	0.0380	-0.7453	0.4626
0	[1, 0.2703]	1	-0.6010	0.0513	-0.5342	0.8326
0	[0.8649, 0.7432]	1	-0.5217	0.0515	-0.4558	0.9460
0	[0.0270, 0.1081]	1	-0.8783	0.0335	-0.8193	0.1644
0	[0.3649, 0.9865]	1	-0.4846	0.0516	-0.4250	0.9223
1	[0, 0.0676]	1	-0.8263	0.0279	-0.8180	0.1897
2	[0.2838, 0.0811]	2	-0.9382	0.0127	-0.8615	0.0946
3	[0.8108, 0.0946]	1	-0.8900	0.0298	-0.8133	0.6218
4	[0.5405, 0.0946]	1	-0.8065	0.0198	-0.8459	0.3516
5	[0.3108, 0.0811]	4	<b>-0.9469</b>	0.0124	-0.8621	<b>0.1216</b>
6	[0.3243, 0.0811]	4	-0.8183	0.0123	-0.8623	0.1351
7	[0, 0]	1	-0.6408	0.0452	-0.6290	0.2058
8	[1, 0.0676]	1	-0.7926	0.0359	-0.7866	0.8109
9	[0.3243, 0.0946]	3	-0.8624	0.0160	-0.8568	0.1358
10	[0.3378, 0.0811]	<b>6</b>	-0.8769	<b>0.0123</b>	<b>-0.8623</b>	0.1486

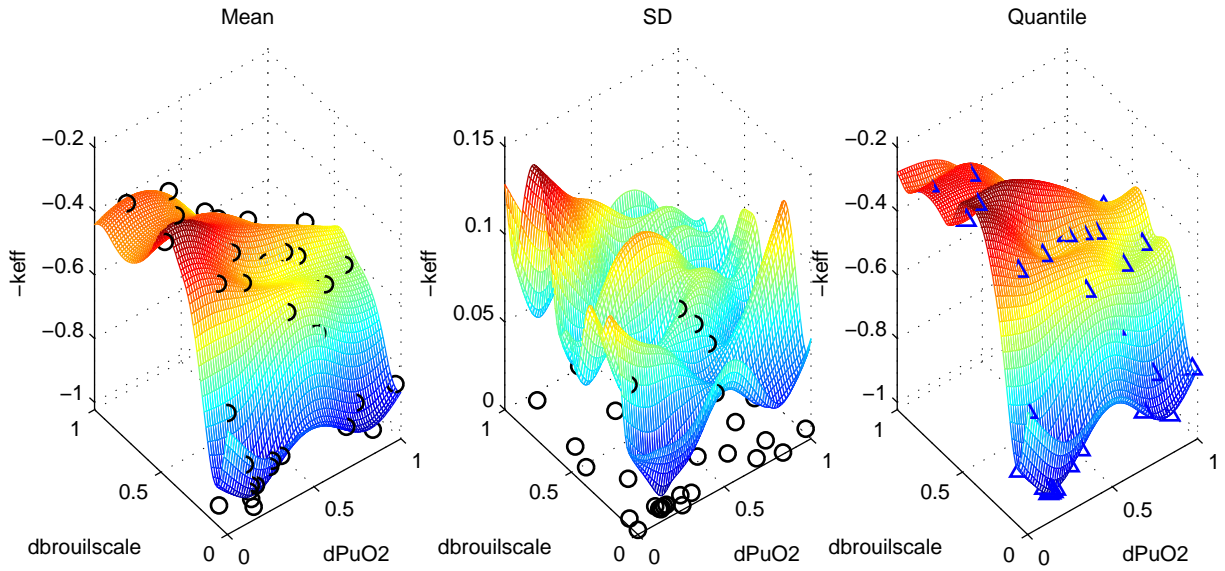


Figure 7: Final kriging after optimization for a  $T = 100$  budget.

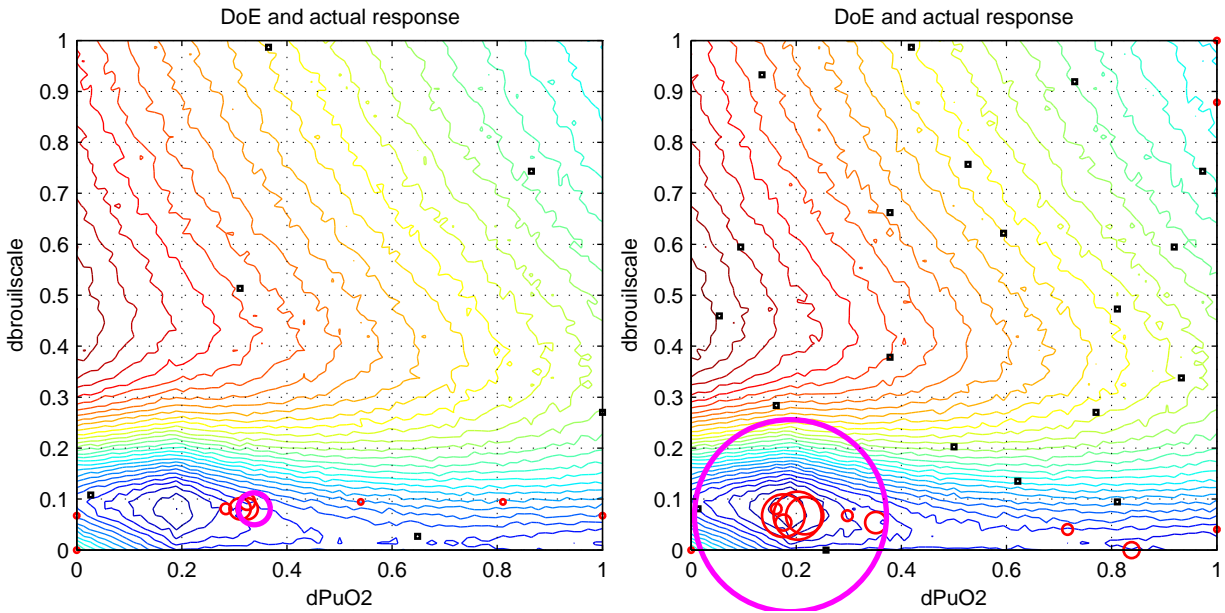


Figure 8: Final DoEs for the budgets  $T = 30$  (left) and  $T = 100$  (right) and contour lines of the actual function. Squares represent the initial measurements and circles the added ones; the markers are proportional to computational time.

Table 5: Sequential measurements for a  $T = 100$  budget.

Iteration	$\mathbf{x}$	Time allocated	$\tilde{y}$	Kriging SD	Kriging quantile	Distance to actual optimum
0	[0.8108, 0.4730]	1	-0.4716	0.0493	-0.4237	0.7348
0	[0.8108, 0.0946]	1	-0.8900	0.0423	-0.8013	0.6218
0	[0.4189, 0.9865]	1	-0.4578	0.0492	-0.4078	0.9341
0	[0.5946, 0.6216]	1	-0.4578	0.0471	-0.4029	0.6757
0	[0.0946, 0.5946]	1	-0.3653	0.0492	-0.2980	0.5222
0	[0.5270, 0.7568]	1	-0.4864	0.0488	-0.4277	0.7554
0	[0.7703, 0.2703]	1	-0.5750	0.0471	-0.5147	0.6111
0	[0.0541, 0.4595]	1	-0.2170	0.0496	-0.1818	0.4018
0	[0.3784, 0.3784]	1	-0.4948	0.0493	-0.4096	0.3524
0	[0.7297, 0.9189]	1	-0.7182	0.0481	-0.6464	0.9971
0	[0.5000, 0.2027]	1	-0.5544	0.0460	-0.4884	0.3338
0	[0.0135, 0.0811]	1	-0.9669	0.0410	-0.8600	0.1757
0	[0.9730, 0.7432]	1	-0.7334	0.0492	-0.6615	1.0260
0	[0.9189, 0.5946]	1	-0.6349	0.0486	-0.5605	0.8923
0	[0.1351, 0.9324]	1	-0.3888	0.0501	-0.3343	0.8531
0	[0.2568, 0 ]	1	-0.8683	0.0418	-0.8039	0.1055
0	[0.9324, 0.3378]	1	-0.5889	0.0484	-0.5247	0.7863
0	[0.6216, 0.1351]	1	-0.6300	0.0422	-0.6176	0.4358
0	[0.3784, 0.6622]	1	-0.3829	0.0465	-0.3337	0.6111
0	[0.1622, 0.2838]	1	-0.3948	0.0495	-0.3562	0.2045
1	[0, 0 ]	1	-0.6408	0.0460	-0.6388	0.2058
2	[0.2162, 0.0676]	7	-0.9603	0.0086	-0.9722	0.0302
3	[0.2973, 0.0676]	2	-0.9260	0.0184	-0.9234	0.1089
4	[0.1622, 0.0811]	2	-0.8837	0.0159	-0.9579	0.0270
5	[0.7162, 0.0405]	2	-0.9159	0.0324	-0.8696	0.5286
6	[0.3514, 0.0541]	4	-0.9027	0.0231	-0.8837	0.1644
7	[1, 0.0405 ]	1	-0.8678	0.0465	-0.8196	0.8118
8	[0.2027, 0.0676]	9	-1.0230	0.0074	-0.9754	0.0191
9	[0.1757, 0.0541]	3	-0.9426	0.0142	-0.9543	0.0302
10	[ 1, 1 ]	1	-0.8610	0.0499	-0.7778	1.2255
11	[ 1, 0.8784 ]	1	-0.7817	0.0480	-0.7240	1.1371
12	[0.1892, 0.0676]	<b>36</b>	-0.9949	<b>0.0071</b>	<b>-0.9760</b>	<b>0.0135</b>
13	[0.1757, 0.0676]	8	-0.9674	0.0080	-0.9735	0.0191
14	[ 0.8378, 0 ]	3	-0.9456	0.0298	-0.9000	0.6537



## 8. CONCLUSION AND PERSPECTIVES

In this paper, we have proposed a quantile-based expected improvement for the optimization of noisy black-box simulators. This criterion allows a rigorous treatment of heterogeneous noise and takes into account the noise level of the candidate measurements. In the context of simulators with tunable fidelity, we proposed an on-line procedure for an adapted distribution of the computational effort. One of the advantages of such procedure is that it prevents from allocating too much time to poor designs, and allows spending more credit on the best ones. Another remarkable property of this algorithm is that, unlike EGO, it takes into account the limited computational budget. Indeed, the algorithm is more exploratory when there is much budget left, and favours a more local search when running out of computational credit. The online allocation optimization algorithm was tested on two problems: an analytical function, and an original application in nuclear criticality safety, the Monte Carlo criticality simulator MORET5. On both problems, the algorithm showed promising results, using coarse measurements for exploration and accurate measurements at best designs. Future work may include comparison of the quantile-based EI to other criteria for point selection, analysis of the effect of on-line allocation compared to a uniform allocation strategy, and a comparison of our algorithm to classical noisy optimization algorithms.

### Acknowledgements

This work was partially supported by French National Research Agency (ANR) through COSINUS program (project OMD2 ANR-08-COSI-007).

## REFERENCES

- Alexandrov, N., Lewis, R., Gumbert, C., Green, L., and Newman, P. (2000), "Optimization with variable-fidelity models applied to wing design," *AIAA paper*, 841(2000), 254.
- Fedorov, V., and Hackl, P. (1997), *Model-oriented design of experiments* Springer.
- Fernex, F., Heulers, L., Jacquet, O., Miss, J., and Richet, Y. (2005), The MORET 4B Monte Carlo code - New features to treat complex criticality systems., in *M&C International Conference on Mathematics and Computation Supercomputing, Reactor Physics and Nuclear and Biological Application, Avignon, France*.
- Forrester, A., Bressloff, N., and Keane, A. (2006), "Optimization using surrogate models and partially converged computational fluid dynamics simulations," *Proceedings of the Royal Society A*, 462(2071), 2177.
- Forrester, A., Keane, A., and Bressloff, N. (2006), "Design and Analysis of Noisy Computer Experiments," *AIAA journal*, 44(10), 2331.

- Gano, S., Renaud, J., Martin, J., and Simpson, T. (2006), “Update strategies for kriging models used in variable fidelity optimization,” *Structural and Multidisciplinary Optimization*, 32(4), 287–298.
- Ginsbourger, D. (2009), Métamodèles Multiples pour l’Approximation et l’Optimisation de Fonctions Numériques Multivariées, PhD thesis, Ecole Nationale Supérieure des Mines de Saint Etienne.
- Ginsbourger, D., and Le Riche, R. (2009), “Towards GP-based optimization with finite time horizon,” <http://hal.archives-ouvertes.fr/hal-00424309/en/>.
- Ginsbourger, D., Picheny, V., Roustant, O., and Richet, Y. (2008), Kriging with Heterogeneous Nugget Effect for the Approximation of Noisy Simulators with Tunable Fidelity,, in *Congrès conjoint de la Société Statistique du Canada et de la SFdS, May 25th-29th, Ottawa (Canada)*.
- Huang, D., Allen, T., Notz, W., and Miller, R. (2006), “Sequential kriging optimization using multiple-fidelity evaluations,” *Structural and Multidisciplinary Optimization*, 32(5), 369–382.
- Huang, D., Allen, T., Notz, W., and Zeng, N. (2006), “Global optimization of stochastic black-box systems via sequential kriging meta-models,” *Journal of Global Optimization*, 34(3), 441–466.
- Jones, D. (2001), “A taxonomy of global optimization methods based on response surfaces,” *Journal of Global Optimization*, 21(4), 345–383.
- Jones, D., Schonlau, M., and Welch, W. (1998), “Efficient global optimization of expensive black-box functions,” *Journal of Global Optimization*, 13(4), 455–492.
- Kennedy, M., and O’Hagan, A. (2000), “Predicting the output from a complex computer code when fast approximations are available,” *Biometrika*, 87(1), 1.
- Marrel, A. (2008), Mise en oeuvre et utilisation du metamodelle processus gaussien pour l’analyse de sensibilité de modeles numeriques: Application a un code de transport hydrogeologique, PhD thesis, INSA Toulouse.
- Matheron, G. (1969), “Le krigeage universel,” *Cahiers du centre de morphologie mathématique*, 1.
- Mockus, J. (1988), *Bayesian Approach to Global Optimization* Kluwer academic publishers.
- Rasmussen, C., and Williams, C. (2006), *Gaussian processes for machine learning* Springer.
- Roustant, O., Ginsbourger, D., and Deville, Y. (2009), “The DiceKriging package: kriging-based metamodelling and optimization for computer experiments,” *Book of abstract of the R User Conference*, .
- Vazquez, E., Villemonteix, J., Sidorkiewicz, M., and Walter, É. (2008), Global optimization based on noisy evaluations: an empirical study of two statistical approaches,, in *Journal of Physics: Conference Series*, Vol. 135, p. 012100.



APPENDIX: ONE-STEP AHEAD CONDITIONAL DISTRIBUTIONS OF THE MEAN, VARIANCE  
AND QUANTILE PROCESSES

Let  $\mathbf{x}^{n+1}$  be the point to be visited at the  $(n+1)^{\text{th}}$  step,  $\tau_{n+1}^2$  and  $\tilde{Y}_{n+1} = Y(\mathbf{x}^{n+1}) + \varepsilon_{n+1}$  the corresponding noise variance and noisy response, respectively. We will now discuss the properties of the Kriging mean and variance at step  $n+1$  seen from step  $n$ . Let  $M_{n+1}(\mathbf{x}) := \mathbb{E}[Y(\mathbf{x})|\tilde{A}_n, \tilde{Y}_{n+1}]$  be the kriging mean function at the  $(n+1)^{\text{th}}$  step and  $S_{n+1}^2(\mathbf{x}) := \text{Var}[Y(\mathbf{x})|\tilde{A}_n, \tilde{Y}_{n+1}]$  the corresponding conditional variance. Seen from step  $n$ , both of them are *ex ante* random processes since they are depending on the not yet observed measurement  $\tilde{Y}_{n+1}$ . We will now prove that they are in fact Gaussian Processes  $|\tilde{A}_n$ , as well as the associated quantile  $Q_{n+1}(\mathbf{x}) = M_{n+1}(\mathbf{x}) + \Phi^{-1}(\beta)S_{n+1}(\mathbf{x})$ . The key results are that the Kriging predictor is linear in the observations, and that the Kriging variance is independent of them, as can be seen from Eqs. 4 and 4. Writing

$$M_{n+1}(\mathbf{x}) = \left( \sum_{j=1}^n \lambda_{n+1,j}(\mathbf{x}) \tilde{Y}_j \right) + \lambda_{n+1,n+1}(\mathbf{x})(Y(\mathbf{x}^{n+1}) + \varepsilon_{n+1}), \text{ where} \quad (17)$$

$$(\lambda_{n+1,\cdot}(\mathbf{x})) := \left( \mathbf{k}_{n+1}(\mathbf{x})^T + \frac{(1 - \mathbf{k}_{n+1}(\mathbf{x})^T (K_{n+1} + \Delta_{n+1})^{-1} \mathbf{1}_{n+1})}{\mathbf{1}_{n+1}^T (K_{n+1} + \Delta_{n+1})^{-1} \mathbf{1}_{n+1}} \mathbf{1}_{n+1}^T \right) (K_{n+1} + \Delta_{n+1})^{-1}, \quad (18)$$

it appears that  $M_{n+1}$  is a GP  $|\tilde{A}_n$ , with the following conditional mean and covariance kernel:

$$\mathbb{E}[M_{n+1}(\mathbf{x})|\tilde{A}_n] = \sum_{j=1}^n \lambda_{n+1,j}(\mathbf{x}) \tilde{y}_j + \lambda_{n+1,n+1}(\mathbf{x}) m_n(\mathbf{x}) \text{ and} \quad (19)$$

$$\text{Cov}[M_{n+1}(\mathbf{x}), M_{n+1}(\mathbf{x}')|\tilde{A}_n] = \lambda_{n+1,n+1}(\mathbf{x}) \lambda_{n+1,n+1}(\mathbf{x}') (s_n^2(\mathbf{x}^{n+1}) + \tau_{n+1}^2). \quad (20)$$

Using that  $Q_{n+1}(\mathbf{x}) = M_{n+1}(\mathbf{x}) + \Phi^{-1}(\beta)S_{n+1}(\mathbf{x})$ , we observe that seen from the  $n^{\text{th}}$  step,  $Q_{n+1}(\cdot)$  is a GP as sum of a GP and a deterministic process conditional on  $\tilde{A}_n$ . We finally get:

$$\mathbb{E}[Q_{n+1}(\mathbf{x})|\tilde{A}_n] = \sum_{j=1}^n \lambda_{n+1,j}(\mathbf{x}) \tilde{y}_j + \lambda_{n+1,n+1}(\mathbf{x}) m_n(\mathbf{x}) + \Phi^{-1}(\beta) s_{n+1}(\mathbf{x}), \quad (21)$$

$$\text{Cov}[Q_{n+1}(\mathbf{x}), Q_{n+1}(\mathbf{x}')|\tilde{A}_n] = \lambda_{n+1,n+1}(\mathbf{x}) \lambda_{n+1,n+1}(\mathbf{x}') (s_n^2(\mathbf{x}^{n+1}) + \tau_{n+1}^2), \quad (22)$$

and the values used in the quantile Expected Improvement (equation 11) are:

$$m_{Q_{n+1}} = \mathbb{E}[Q_{n+1}(\mathbf{x}^{n+1})|\tilde{A}_n] \quad (23)$$

$$s_{Q_{n+1}}^2 = \text{Var} [Q_{n+1}(\mathbf{x}^{n+1})|\tilde{A}_n] = (\lambda_{n+1,n+1}(\mathbf{x}^{n+1}))^2 (s_n^2(\mathbf{x}^{n+1}) + \tau_{n+1}^2) \quad (24)$$