



HAL
open science

Adaptive Design of Experiments for Conservative Estimation of Excursion Sets

Dario Azzimonti, David Ginsbourger, Clément Chevalier, Julien Bect, Yann
Richet

► **To cite this version:**

Dario Azzimonti, David Ginsbourger, Clément Chevalier, Julien Bect, Yann Richet. Adaptive Design of Experiments for Conservative Estimation of Excursion Sets. 2016. hal-01379642v1

HAL Id: hal-01379642

<https://hal.science/hal-01379642v1>

Preprint submitted on 11 Oct 2016 (v1), last revised 28 Jan 2020 (v6)

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.

Adaptive Design of Experiments for Conservative Estimation of Excursion Sets

Dario Azzimonti*, David Ginsbourger[†]*, Clément Chevalier[‡],
Julien Bect[§], Yann Richet[¶]

Abstract

We consider a Gaussian process model trained on few evaluations of an expensive-to-evaluate deterministic function and we study the problem of estimating a fixed excursion set of this function. We review the concept of conservative estimates, recently introduced in this framework, and, in particular, we focus on estimates based on Vorob'ev quantiles. We present a method that sequentially selects new evaluations of the function in order to reduce the uncertainty on such estimates. The sequential strategies are first benchmarked on artificial test cases generated from Gaussian process realizations in two and five dimensions, and then applied to two reliability engineering test cases.

1 Introduction

The problem of estimating the set of inputs that leads a system to a particular behaviour is common in many applications, notably reliability engineering (see, e.g., Bect et al., 2012; Chevalier et al., 2014a), climatology (see, e.g., French and Sain, 2013; Bolin and Lindgren, 2015) and many other fields (see, e.g., Bayarri et al., 2009; Arnaud et al., 2010; Wheeler et al., 2014). Here we consider a system modelled as a continuous function $f : \mathbb{X} \rightarrow \mathbb{Y}$, where \mathbb{X} is

*IMSV, Department of Mathematics and Statistics, University of Bern, Alpeneggstrasse 22, 3012 Bern, Switzerland.

[†]Uncertainty Quantification and Optimal Design group, Idiap Research Institute, Centre du Parc, Rue Marconi 19, PO Box 592, 1920 Martigny, Switzerland.

[‡]Institute of Statistics, University of Neuchâtel, Avenue de Bellevaux 51, 2000 Neuchâtel, Switzerland.

[§]Laboratoire des Signaux et Systèmes (UMR CNRS 8506), CentraleSupélec, CNRS, Univ Paris-Sud, Université Paris-Saclay, 91192, Gif-sur-Yvette, France.

[¶]Institut de Radioprotection et de Sûreté Nucléaire (IRSN), Paris, France.

a locally compact Hausdorff topological space and $\mathbb{Y} = \mathbb{R}^k$, for some $k > 0$. We are interested in the set

$$\Gamma^* = \{x \in \mathbb{X} : f(x) \in T\}, \quad (1)$$

where $T \subset \mathbb{Y}$ is closed, which implies that Γ^* is closed as f is continuous. We focus on the case where few evaluations of the response f can be made available. This is common if f is expensive-to-evaluate, as, for example, when its evaluations require time consuming computer experiments (Sacks et al., 1989).

In a Bayesian framework (see, e.g., Chilès and Delfiner, 2012, and references therein) we assume that f is a realization of an almost surely continuous Gaussian process (GP) $Z \sim GP(\mathbf{m}, \mathfrak{K})$, with mean function \mathbf{m} , defined as $\mathbf{m}(x) := \mathbb{E}[Z_x]$, $x \in \mathbb{X}$, and covariance kernel $\mathfrak{K}(x, y)$, defined as $\mathfrak{K}(x, y) := \text{Cov}(Z_x, Z_y)$, $x, y \in \mathbb{X}$. For $n > 0$, we consider evaluations $\mathbf{f}_n = (f(x_1), \dots, f(x_n))$ of f at an initial design $\mathbf{X}_n = (x_1, \dots, x_n) \in \mathbb{X}^n$ and the posterior distribution of the process $Z \mid Z_{\mathbf{X}_n} = \mathbf{f}_n$, where $Z_{\mathbf{X}_n} = (Z_{x_1}, \dots, Z_{x_n})$. The prior distribution on Z induces a prior distribution on the excursion set

$$\Gamma = \{x \in \mathbb{X} : Z_x \in T\}. \quad (2)$$

Analogously, the posterior distribution on Z induces a posterior distribution for Γ and by summarizing this distribution we obtain an estimate for Γ^* . Random closed sets do not have a unique definition of expectation, however it is possible to provide different notions of expectation and variability for such objects (see, e.g., Molchanov, 2005). The Vorob'ev expectation (Vorob'ev, 1984) is an example recently introduced in the GP framework (see, e.g., Chevalier et al., 2013). Here we focus on conservative estimates for Γ introduced by French and Sain (2013); Bolin and Lindgren (2015).

Consider a Borel σ -finite measure μ defined on \mathbb{X} and denote with \mathfrak{C} a family of closed subsets in \mathbb{X} . A *conservative estimate at level α* for Γ^* is a set $\text{CE}_{\alpha, n}$ defined as

$$\text{CE}_{\alpha, n} \in \arg \max_{C \in \mathfrak{C}} \{\mu(C) : P_n(C \subset \Gamma) \geq \alpha\}, \quad (3)$$

where $P_n(\cdot) = P(\cdot \mid Z_{\mathbf{X}_n} = \mathbf{f}_n)$. This type of set estimate is particularly interesting in problems where Γ^* is a set of safe configurations for a system. In this case, by choosing a high α , we provide an estimate for Γ^* that with high posterior probability is included in Γ . The optimization procedure in equation (3) can be very challenging to solve, and it crucially depend on the choice of the family \mathfrak{C} . A very common choice (see, e.g., French and Sain, 2013; Bolin and Lindgren, 2015; Azzimonti and Ginsbourger, 2016) is a

parametric family of nested sets depending on a one dimensional parameter. Here we also rely on a specific one dimensional family and, in Section 2, we provide motivation for this choice.

The main contribution of this work is a framework to evaluate the uncertainty on conservative estimates for Γ^* and the development of strategies to sequentially reduce such uncertainty by adding new evaluations of f . If the input space \mathbb{X} is a compact subset of \mathbb{R}^d and the output space $\mathbb{Y} = \mathbb{R}$, sequential strategies have already been proposed for contour lines (Ranjan et al., 2008; Bichon et al., 2008) by adapting the expected improvement algorithm criterion (Jones et al., 1998). In the case of excursion sets, Stepwise Uncertainty Reduction (SUR) strategies based on the set’s measure were introduced by Vazquez and Bect (2009); Bect et al. (2012). More recently a fast parallel implementation of these strategies have been proposed (Chevalier et al., 2014a) and applied to the problem of identifying the set Γ^* . Here we extend this framework to conservative estimates.

In Section 1.1 we briefly recall the set estimates preliminary to this work and, in Section 1.2, the previously introduced uncertainty reduction techniques. In Section 2 we study the conservative estimates and we motivate our choice for the family \mathfrak{C} . In Section 3 we define the metrics used to quantify the uncertainty on conservative estimates. In Section 4, we detail the proposed sequential strategies and their implementation in real valued case ($\mathbb{Y} = \mathbb{R}$). In Section 5 we first benchmark the strategies with Gaussian process realizations showing how the conservative property affects the estimates. We then apply the proposed strategies on two industrial test cases. The first one is a coastal flood problem and the second one is a nuclear criticality safety problem.

1.1 The Vorob’ev approach to excursion set estimation

The posterior distribution of Γ provides estimates for Γ^* and different ways to quantify the uncertainty on these estimates. See, e.g. Chevalier et al. (2014a); Bolin and Lindgren (2015); Azzimonti et al. (2016) for more details.

Let us now briefly recall the Vorob’ev approach (Vorob’ev, 1984; Molchanov, 2005; Chevalier et al., 2013). Define the *coverage probability function* of a random closed set Γ as

$$p_{\Gamma}(x) = P(x \in \Gamma), \quad x \in \mathbb{X}.$$

In our case we consider the posterior coverage function $p_{\Gamma,n}$, where we consider the posterior probability P_n . The coverage function defines the family of *Vorob’ev quantiles*

$$Q_{n,\rho} = \{x \in \mathbb{X} : p_{\Gamma,n}(x) \geq \rho\}, \quad (4)$$

with $\rho \in [0, 1]$. These sets are closed for each $\rho \in [0, 1]$ as the coverage function is upper semi-continuous (Molchanov, 2005, Proposition 1.34).

The level ρ can be selected in different ways. For example by choosing $\rho = 0.5$ we obtain the *Vorob'ev median*. Another important type of set estimate based on the Vorob'ev quantiles is the *Vorob'ev expectation* of Γ , where the level $\rho = \rho_V$ is chosen such that Q_ρ has the closest possible measure to the expected measure of Γ , for some Borel σ -finite measure on \mathbb{X} . In Section 2 we show that Vorob'ev quantiles are a reasonable choice for the family \mathfrak{C} in the conservative estimates defined in Equation (3).

Finally let us introduce the concept of expected distance in measure between two random sets. Consider two random closed sets $\Gamma_1, \Gamma_2 \subset \mathbb{X}$. The *expected distance in measure* between Γ_1 and Γ_2 with respect to the measure μ is

$$d_\mu(\Gamma_1, \Gamma_2) = \mathbb{E}[\mu(\Gamma_1 \Delta \Gamma_2)],$$

where $\Gamma_1 \Delta \Gamma_2 = (\Gamma_1 \setminus \Gamma_2) \cup (\Gamma_2 \setminus \Gamma_1)$. The expected distance in measure can be seen as the measure of variability associated with a Vorob'ev quantile and with the Vorob'ev expectation. In the Ph.D. thesis of Chevalier (2013), this notion of variability was introduced to adaptively reduce the uncertainty on Vorob'ev expectations for expensive-to-evaluate functions. Here, it is used in Section 3 to provide uncertainty functions for conservative estimates.

1.2 Background on SUR strategies

The objective of sequential strategies for GP models is to select a sequence of points X_1, X_2, \dots, X_n that reduces the uncertainty on selected posterior quantities. Here we are interested in reducing the uncertainties on $\text{CE}_{\alpha, n}$, as defined in equation (3). In this case the current state of the model is described by $\mathbf{X}_n \in \mathbb{X}^n$ and $Z_{\mathbf{X}_n} \in \mathbb{Y}^n$, with \mathbf{X}_n the locations where the function f was evaluated and $Z_{\mathbf{X}_n}$ are the actual evaluations. Here, we focus on Stepwise Uncertainty Reduction (SUR) strategies. A SUR strategy (see, e.g., Fleuret and Geman, 1999; Bect et al., 2012; Chevalier et al., 2014a) selects the next evaluation in order to reduce a particular uncertainty function.

Definition 1 (Uncertainty function). Consider a model where n evaluations of f are given, we call uncertainty of an estimate the map

$$H_n : (\mathbb{X} \times \mathbb{Y})^n \rightarrow \mathbb{R},$$

that associates to each vector of couples $(x_i, Z_{x_i})_{i=1, \dots, n}$ a real value representing the uncertainty associated with the selected estimate.

Since they are selected sequentially, both the locations X_1, \dots, X_n and the evaluations are random. More specifically, denote with \mathcal{A}_n the σ -algebra generated by the couples $X_1, Z_{X_1}, \dots, X_n, Z_{X_n}$. We denote with \mathcal{H}_n the \mathcal{A}_n measurable random variable that returns the uncertainty associated with \mathcal{A}_n , the σ -algebra generated by $(X_i, Z_{X_i})_{i=1, \dots, n}$.

In what follows we denote with $\mathbb{E}_n[\cdot] = \mathbb{E}[\cdot \mid Z_{\mathbf{X}_n} = \mathbf{f}_n]$, the expectation conditioned on the event $Z_{\mathbf{X}_n} = \mathbf{f}_n$, where $\mathbf{X}_n \in \mathbb{X}^n$ is a fixed design and $\mathbf{f}_n \in \mathbb{Y}^n$. If we assume that the first n evaluations of the field are known then, a SUR strategy selects the locations $X_{n+1}^*, \dots, X_{n+q}^*$ that minimize $\mathbb{E}_n[\mathcal{H}_{n+q}]$, the future uncertainty in expectation. For a more complete and theoretical perspective on SUR strategies see, e.g., Bect et al. (2016) and references therein. There are many ways to proceed with the minimization introduced above, see, e.g., Osborne et al. (2009); Ginsbourger and Le Riche (2010); Bect et al. (2012); González et al. (2016) and references therein. Here we focus on sub-optimal strategies, also called batch-sequential *one-step lookahead strategies*, that select the next batch of locations by greedily minimizing the expected uncertainty at the next step. This choice is often justified in practice because it is possible to run the evaluations of the function in parallel thus saving wall-clock time.

Definition 2 (batch sequential one-step lookahead criterion). We call batch sequential one-step lookahead sampling criterion a function $J_n : (\mathbb{X})^q \rightarrow \mathbb{R}$ that associates to each batch of q points $\mathbf{x}^{(q)} := (x_{n+1}, \dots, x_{n+q}) \in \mathbb{X}^q$ the expected uncertainty at the next step assuming this batch is evaluated

$$J_n(\mathbf{x}^{(q)}) = \mathbb{E}_n[\mathcal{H}_{n+q} \mid X_{n+1} = x_{n+1}, \dots, X_{n+q} = x_{n+q}].$$

In the next sections we revisit these concepts for the problem of computing conservative estimates of an excursion set.

2 Properties of conservative estimate

The conservative estimate introduced in Equation (3) requires the specification of the family \mathfrak{C} where to search for the optimal set CE_α . In practice, it is convenient to choose a parametric family indexed by a parameter $\theta \in \mathbb{R}^k$. Consider a nested family \mathfrak{C}_θ indexed by a real number $\theta \in [0, 1]$, i.e. for each $\theta_1 > \theta_2$

$$C_{\theta_1} \subset C_{\theta_2}, \tag{5}$$

for any $C_{\theta_1}, C_{\theta_2} \in \mathfrak{C}_\theta$. Let us define $C_0 = \mathbb{X}$ and assume that $\mu(\mathbb{X}) < \infty$. In practice, this is often the case as \mathbb{X} is either chosen as a compact subset of \mathbb{R}^d with μ the Lebesgue measure or μ is a probability measure.

For each θ , we can define the function $\varphi_\mu : [0, 1] \rightarrow [0, +\infty)$ that associates to each $\theta \in [0, 1]$ the value $\varphi_\mu(\theta) := \mu(C_\theta)$, with $C_\theta \in \mathfrak{C}_\theta$. It is a non increasing function of θ because the sets C_θ are nested. We further define the function $\psi_\Gamma : [0, 1] \rightarrow [0, 1]$ that associates to each θ the probability $\psi_\Gamma(\theta) := P(C_\theta \subset \Gamma)$. The function ψ_Γ is non decreasing in θ due to the nested property in Equation (5). In this set-up the computation of CE_α is reduced to finding the smallest $\theta = \theta^*$ such that $\psi_\Gamma(\theta^*) \geq \alpha$. The measure $\mu(C_{\theta^*})$ is equal to $\varphi_\mu(\theta^*)$.

The Vorob'ev quantiles introduced in Equation (4) are a family of nested closed sets that satisfy the property in Equation (5). Moreover they have the important property of being the minimizers of the expected distance in measure among sets with the same measure.

Proposition 1. *Consider a measure μ such that $\mu(\mathbb{X}) < \infty$. The Vorob'ev quantile*

$$Q_\rho = \{x \in \mathbb{X} : p_\Gamma(x) \geq \rho\}$$

minimizes the expected distance in measure with Γ among measurable sets M such that $\mu(M) = \mu(Q_\rho)$, i.e.,

$$\mathbb{E}[\mu(Q_\rho \Delta \Gamma)] \leq \mathbb{E}[\mu(M \Delta \Gamma)], \quad (6)$$

for each measurable set M such that $\mu(M) = \mu(Q_\rho)$.

Proof. see Appendix A □

The Vorob'ev quantiles have the smallest expected distance in measure with respect to Γ among sets with the same measure. They are thus an optimal family for conservative estimates with respect to the expected distance in measure. In general, however, the Vorob'ev quantile chosen for CE_α with this procedure is not the set S with the largest measure satisfying the property $P(S \subset \Gamma) \geq \alpha$.

In the remainder of the paper we always consider \mathfrak{C} as the family of Vorob'ev quantiles. Given an initial design \mathbf{X}_n we can exploit the previously described properties and obtain $\text{CE}_{\alpha,n}$, a conservative estimate at level α for Γ^* . In the next section we introduce different ways to quantify the uncertainty on this estimate, while in Section 4 we introduce sequential strategies to reduce this uncertainty by adding new evaluations to the model.

3 Uncertainty quantification on $\text{CE}_{\alpha,n}$

In this section we introduce several uncertainty functions for conservative estimates. Here we consider a static scenario where n evaluations of f are available.

Our object of interest is Γ^* , therefore we require uncertainty functions that take into account the whole set. Chevalier and co-authors (Chevalier et al., 2013; Chevalier, 2013) evaluate the uncertainty on the Vorob'ev expectation with the expected distance in measure between the current estimate Q_{n,ρ_n} and the set Γ . Let us recall here that the *Vorob'ev uncertainty* of the quantile Q_{ρ_n} is the quantity

$$\mathcal{H}_n(\rho_n) = \mathbb{E}_n[\mu(\Gamma \Delta Q_{n,\rho_n})]. \quad (7)$$

In the following sections, this uncertainty quantification function is applied to the Vorob'ev expectation, $\rho_n = \rho_{V,n}$, to the Vorob'ev median, $\rho_n = 0.5$, and to the conservative estimate at level α , $\rho_n = \rho_n^\alpha$.

Consider now the case of conservative estimates and fix the level α . Note that, by definition, the symmetric difference can be written as

$$\mathbb{E}_n[\mu(\Gamma \Delta Q_{n,\rho_n^\alpha})] = \mathbb{E}_n[\mu(Q_{n,\rho_n^\alpha} \setminus \Gamma)] + \mathbb{E}_n[\mu(\Gamma \setminus Q_{n,\rho_n^\alpha})]. \quad (8)$$

Let us denote with $G_n^{(1)} = \mu(Q_{n,\rho_n^\alpha} \setminus \Gamma)$ the random variable associated with the measure of the first set difference and with $G_n^{(2)} = \mu(\Gamma \setminus Q_{n,\rho_n^\alpha})$ the random variable associated with the second one.

Remark 1. Consider the conservative estimate Q_{n,ρ_n^α} , then the ratio between the error $\mathbb{E}_n[G_n^{(1)}]$ and the measure $\mu(Q_{n,\rho_n^\alpha})$ is bounded by $1 - \alpha$, the chosen level for the conservative estimates. See Appendix A for a proof.

A conservative estimate Q_{n,ρ_n^α} aims at controlling the error $\mathbb{E}_n[G_n^{(1)}]$. With a broad use of the hypothesis testing lexicon we denote *Type I* error at state n the quantity $\mathbb{E}_n[G_n^{(1)}]$ and *Type II* error at state n the quantity $\mathbb{E}_n[G_n^{(2)}]$. Type II error defines the following uncertainty function for $\text{CE}_{\alpha,n}$.

Definition 3 (Type II uncertainty). Consider the Vorob'ev quantile Q_{n,ρ_n^α} corresponding to the conservative estimate at level α for Γ . The *Type II uncertainty* is the uncertainty function $\mathcal{H}_n^{\text{T2}}$ defined as

$$\mathcal{H}_n^{\text{T2}}(\rho_n^\alpha) := \mathbb{E}_n[G_n^{(2)}] = \mathbb{E}_n[\mu(\Gamma \setminus Q_{n,\rho_n^\alpha})] \quad (9)$$

Conservative estimates at high levels α tend to select regions inside Γ , by definition. In particular if the number of function evaluations is high enough to have a good approximation of the function f , the conservative estimates with high α tend to be inside the true excursion set Γ^* . In these situations the expected type I error is usually very small, as shown in Remark 1, while type II error could be rather large. Type II uncertainty is thus a relevant

quantity when evaluating conservative estimates. In the test case studies we also compute the expected type I error to check that it is consistently small.

The last uncertainty function introduced in this section is the measure of Q_{n,ρ_n^α} multiplied by -1 . In fact by definition the conservative estimate has probability at least α of being inside the excursion. If an estimate Q_{n,ρ_n^α} has a large measure it is then a good indicator of how well the target set is covered by the estimate while still retaining the conservative statement.

Definition 4 (Uncertainty MEAS). We denote the uncertainty function related to the measure μ with $\mathcal{H}_n^{\text{MEAS}}$, defined as

$$\mathcal{H}_n^{\text{MEAS}}(\rho_n^\alpha) := -\mathbb{E}_n[\mu(Q_{n,\rho_n^\alpha})] \quad (10)$$

If $\mu(\mathbb{X}) < \infty$, then $\mathcal{H}_n^{\text{MEAS}}(\rho_n^\alpha)$ is bounded from below by $-\mu(\mathbb{X})$. While the uncertainty functions are usually bounded from below by zero, this property is not required to develop SUR strategies.

4 SUR strategies for conservative estimates

In this section we consider a current design of experiments \mathbf{X}_n , for some $n > 0$. We introduce one-step lookahead SUR strategies that select the next batch of $q > 0$ locations $X_{n+1}, \dots, X_{n+q} \in \mathbb{X}$ in order to reduce the future uncertainty \mathcal{H}_{n+q} defined in Equation (7), (9) and (10). We denote with \mathcal{A}_{n+q} the σ -algebra generated by the sequence $(X_i, Z_{X_i})_{i=1,\dots,n+q} \in (\mathbb{X} \times \mathbb{Y})^{n+q}$. In this case the uncertainty \mathcal{H}_{n+q} is a random variable that depends on the unknown response at the points X_{n+1}, \dots, X_{n+q} .

In the case of conservative estimates we consider three sampling criterion based on the uncertainty functions introduced in Equation (7), (9) and (10). The first is an adaptation of the Vorob'ev criterion introduced in Chevalier (2013) and based on the Vorob'ev deviation (Vorob'ev, 1984; Molchanov, 2005; Chevalier et al., 2013).

$$\begin{aligned} J_n(\mathbf{x}^{(q)}; \rho_{n+q}^\alpha) &= \mathbb{E}_n \left[\mathcal{H}_{n+q}(\rho_{n+q}^\alpha) \mid X_{n+1} = x_{n+1}, \dots, X_{n+q} = x_{n+q} \right] \\ &= \mathbb{E}_n \left[\mathbb{E}_{n+q} \left[\mu(\Gamma \Delta Q_{n+q, \rho_{n+q}^\alpha}) \mid X_{n+1} = x_{n+1}, \dots, X_{n+q} = x_{n+q} \right] \right] \end{aligned} \quad (11)$$

for $\mathbf{x}^{(q)} = (x_{n+1}, \dots, x_{n+q}) \in \mathbb{X}^q$, where $Q_{n+q, \rho_{n+q}^\alpha}$ is the Vorob'ev quantile obtained with $n+q$ evaluations of the function at level ρ_{n+q}^α , the conservative level obtained with $n+q$ evaluations. In what follows for the sake of brevity we denote with $\mathcal{A}_{n+q}(\mathbf{x}^{(q)})$ the σ -algebra generated by the couples $(x_{n+1}, Z_{x_{n+1}}), \dots, (x_{n+q}, Z_{x_{n+q}})$, where $(x_{n+1}, \dots, x_{n+q}) = \mathbf{x}^{(q)}$.

In the case of conservative estimates with high level α , each term of Equation (8) does not contribute equally to the expected distance in measure, as observed in Remark 1. It is thus reasonable to consider a criterion to reduce the Type II uncertainty introduced in Definition 3.

$$\begin{aligned} J_n^{\text{T}2}(\mathbf{x}^{(q)}; \rho_{n+q}^\alpha) &= \mathbb{E}_n \left[\mathcal{H}_n^{\text{T}2}(\rho_{n+q}^\alpha) \mid \mathcal{A}_{n+q}(\mathbf{x}^{(q)}) \right] \\ &= \mathbb{E}_n \left[\mathbb{E} \left[G_n^{(2)}(Q_{n+q, \rho_{n+q}^\alpha}) \mid \mathcal{A}_{n+q}(\mathbf{x}^{(q)}) \right] \right], \end{aligned} \quad (12)$$

for $\mathbf{x}^{(q)} \in \mathbb{X}^q$. In Section 4.1 this criterion is derived for a generic level $\rho_n \in [0, 1]$, under more restrictive assumptions on the process and on Γ^* .

The last criterion studied here for conservative estimates relies on the uncertainty function $\mathcal{H}_n^{\text{MEAS}}$. We can define the measure based criterion as

$$J_n^{\text{MEAS}}(\mathbf{x}^{(q)}; \rho_{n+q}^\alpha) = \mathbb{E}_n \left[-\mathbb{E} \left[\mu(Q_{n+q, \rho_{n+q}^\alpha}) \mid \mathcal{A}_{n+q}(\mathbf{x}^{(q)}) \right] \right]. \quad (13)$$

Since we are interested in minimizing this criterion we consider the equivalent function to maximize $\widetilde{J}_n^{\text{MEAS}}(\mathbf{x}^{(q)}; \rho_{n+q}^\alpha) = \mathbb{E}_n \left[\mathbb{E} \left[\mu(Q_{n+q, \rho_{n+q}^\alpha}) \mid \mathcal{A}_{n+q}(\mathbf{x}^{(q)}) \right] \right]$. Note that this criterion select points that are meant to increase the measure of the estimate and it is only reasonable for conservative estimates where the conservative condition on $Q_{\rho_{n+q}^\alpha}$ leads to sets with finite measure in expectation. In practice, there is no closed form formula the criteria presented above with a choice of ρ_{n+q}^α . In the implementation part we replace this level with ρ_n^α , the last level computed.

4.1 Implementation

In this section we detail the algorithmic aspects of the criteria.

The notions of an estimate's uncertainty and of sequential criteria can be defined in the generic setting introduced in Section 1.2, however in order to provide formulae to implement the criteria we need to restrict ourselves to a more specific framework. Here we fix $\mathbb{X} \subset \mathbb{R}^d$, a compact subset of \mathbb{R}^d , and $\mathbb{Y} = \mathbb{R}$. These choices are common especially in engineering and other scientific applications. We assume that Z is a GP with constant prior mean \mathbf{m} and covariance kernel \mathfrak{K} . Finally we derive formulas for the criteria for the set $\Gamma^* = \{x \in \mathbb{X} : f(x) \in T\}$ with $T = [t, +\infty)$, where $t \in \mathbb{R}$ is a fixed threshold. It is straightforward to compute the criteria for $T = (-\infty, t]$ and to extend them for unions of such intervals.

The formulas for the criteria introduced here all rely on the posterior coverage probability function p_n , where the subscript Γ is dropped as the set is clear from the context. In particular, from the assumptions previously

introduced it follows that, for each $n \geq 0$, $p_n(x) = \Phi((\mathbf{m}_n(x) - t)/s_n(x))$, with $x \in \mathbb{X}$, where Φ is the univariate standard Normal distribution, \mathbf{m}_n is the posterior mean of the process and $s_n(x) = \sqrt{\mathfrak{K}(x, x)}$ for all $x \in \mathbb{X}$.

The first criterion introduced in Section 4 is based on the symmetric difference between the set Γ and the Vorob'ev quantile $Q_{n,\rho}$. In Chevalier (2013), Chapter 4.2, the formula for this criterion in this framework was derived for the Vorob'ev expectation, i.e. the quantile at level $\rho = \rho_V$. In the following remark we first extend this result to any quantile ρ .

Remark 2 (Criterion J_n). Under the previously introduced assumptions the criterion J_n can be expanded in closed-form as

$$\begin{aligned} J_n(\mathbf{x}^{(q)}; \rho_n) &= \mathbb{E}_n \left[\mathbb{E} \left[\mu(\Gamma \Delta Q_{n,q,\rho_n}) \mid \mathcal{A}_{n+q}(\mathbf{x}^{(q)}) \right] \right] \\ &= \int_{\mathbb{X}} \left(2\Phi_2 \left(\left(\begin{array}{c} a_{n+q}(u) \\ \Phi^{-1}(\rho_n) - a_{n+q}(u) \end{array} \right); \begin{pmatrix} 1 + \gamma_{n+q}(u) & -\gamma_{n+q}(u) \\ -\gamma_{n+q}(u) & \gamma_{n+q}(u) \end{pmatrix} \right) \right. \\ &\quad \left. - p_n(u) + \Phi \left(\frac{a_{n+q}(u) - \Phi^{-1}(\rho_n)}{\gamma_{n+q}(u)} \right) \right) d\mu(u), \end{aligned} \quad (14)$$

where

$$\begin{aligned} a_{n+q}(u) &= \frac{\mathbf{m}_n(u) - t}{s_{n+q}(u)}, & \mathbf{b}_{n+q}(u) &= \frac{\mathfrak{K}_n(u, \mathbf{x}^{(q)}) K_q^{-1}}{s_{n+q}(u)}, \\ \gamma_{n+q}(u) &= \sqrt{\mathbf{b}_{n+q}^T(u) K_q \mathbf{b}_{n+q}(u)}, \quad u \in \mathbb{X} \end{aligned} \quad (15)$$

with $\mathfrak{K}_n(u, \mathbf{x}^{(q)}) = (\mathfrak{K}(u, x_{n+1}), \dots, \mathfrak{K}(u, x_{n+q}))$. K_q is covariance matrix with elements $[\mathfrak{K}_n(x_{n+i}, x_{n+j})]_{i,j=1,\dots,q}$, $\Phi_2(\cdot; \Sigma)$ is the bivariate centred Normal distribution with covariance matrix Σ and $\Phi^{-1}(u)$ denotes the quantile at level u of the standard Normal distribution.

The proof of the previous remark is a simple adaptation of the proof in Chevalier (2013), Chapter 4.2.

Proposition 2. *The criterion $J_n^{\Gamma 2}(\cdot; \rho_n^\alpha)$ can be expanded in closed-form as*

$$\begin{aligned} J_n^{\Gamma 2}(\mathbf{x}^{(q)}; \rho_n^\alpha) &= \mathbb{E}_n \left[\mathbb{E} \left[G_n^{(2)}(Q_{n+q,\rho_n^\alpha}) \mid \mathcal{A}_{n+q}(\mathbf{x}^{(q)}) \right] \right] \\ &= \int_{\mathbb{X}} \Phi_2 \left(\left(\begin{array}{c} a_{n+q}(u) \\ \Phi^{-1}(\rho_n^\alpha) - a_{n+q}(u) \end{array} \right); \begin{pmatrix} 1 + \gamma_{n+q}(u) & -\gamma_{n+q}(u) \\ -\gamma_{n+q}(u) & \gamma_{n+q}(u) \end{pmatrix} \right) d\mu(u). \end{aligned} \quad (16)$$

Proof. The proof is a simple adaptation of the Remark 2. See Chevalier (2013). \square

Proposition 3. *The criterion J_n^{MEAS} can be expanded in closed-form as*

$$\begin{aligned} J_n^{\text{MEAS}}(\mathbf{x}^{(q)}; \rho_n^\alpha) &= \mathbb{E}_n \left[\mathbb{E} \left[\mu(Q_{n+q,\rho_n^\alpha}) \mid \mathcal{A}_{n+q}(\mathbf{x}^{(q)}) \right] \right] \\ &= \int_{\mathbb{X}} \Phi \left(\frac{a_{n+q}(u) - \Phi^{-1}(\rho_n^\alpha)}{\gamma_{n+q}(u)} \right) d\mu(u). \end{aligned} \quad (17)$$

Proof. First of all notice that, for each $x \in \mathbb{X}$ the coverage function $p_{n+q}(\mathbf{x}^{(q)})$ can be written as

$$p_{n+q}(\mathbf{x}^{(q)})(x) = \Phi \left(a_{n+q}(x) + \mathbf{b}_{n+q}^T Y_q \right), \quad (18)$$

where a_{n+q} , \mathbf{b}_{n+q} are defined in Equation (15) and $Y_q \sim N_q(0, K_q)$ is a q dimensional normal random vector. The indicator function of the set Q_{n+q, ρ_n^α} can be written as $\mathbf{1}_{p_{n+q}(x) \geq \rho_n^\alpha}$. By Tonelli's theorem we exchange the expectation with the integral over \mathbb{X} and we obtain

$$\begin{aligned} \mathbb{E}_n \left[\mathbb{E} \left[\mu(Q_{n+q, \rho_n^\alpha}) \mid \mathcal{A}_{n+q}(\mathbf{x}^{(q)}) \right] \right] &= \int_{\mathbb{X}} \mathbb{E}_n \left[\mathbf{1}_{p_{n+q}(u) \geq \rho_n^\alpha} \right] d\mu(u) \\ &= \int_{\mathbb{X}} P_n(p_{n+q}(u) \geq \rho_n^\alpha) d\mu(u). \end{aligned}$$

By substituting the expression in Equation (18) we obtain

$$\begin{aligned} \int_{\mathbb{X}} P_n(p_{n+q}(u) \geq \rho_n^\alpha) d\mu(u) &= \int_{\mathbb{X}} P_n(a_{n+q}(u) + \mathbf{b}_{n+q}^T(u) Y_q \geq \Phi^{-1}(\rho_n^\alpha)) d\mu(u) \\ &= \int_{\mathbb{X}} \Phi \left(\frac{a_{n+q}(u) - \Phi^{-1}(\rho_n^\alpha)}{\gamma_{n+q}(u)} \right) d\mu(u) \end{aligned}$$

□

For the practical implementation of the sampling criteria we exploit the kriging update formulas (Chevalier et al., 2014b; Emery, 2009) for faster updates of the posterior mean and covariance when new evaluations are added.

The sampling criteria, implemented above in Equation (14), (16) and (17), are used to select the next evaluations of the function f . The conservative level ρ_n^α is computed with the algorithm detailed by Azzimonti and Ginsbourger (2016).

5 Test cases

In this section we apply the proposed sequential uncertainty reduction methods to different test cases. First we develop a benchmark study with Gaussian process realizations to study the different behaviour of the proposed strategies. Then, we apply the methods to two reliability engineering test cases. In the first test case, the set of interest represents the offshore conditions that do not lead the water level at the coast to be larger than a critical threshold above which flood would occur. In the second test case the set of interest is the set of safe parameters for nuclear storage facility.

Strategy number	criterion	parameters
Benchmark 1	IMSE	
Benchmark 2	tIMSE	target= t
A	$J_n(\cdot; \rho_n)$	$\rho_n = 0.5$
B	$J_n(\cdot; \rho_n)$	$\rho_n = \rho_n^\alpha, \alpha = 0.95$
C	$J_n^{\text{MEAS}}(\cdot; \rho_n)$	$\rho_n = \rho_n^\alpha, \alpha = 0.95$
D	$J_n^{\text{T}2}(\cdot; \rho_n)$	$\rho_n = 0.5$
E	$J_n^{\text{T}2}(\cdot; \rho_n^\alpha)$	$\alpha = 0.95$
F (hybrid strategy)	$J_n^{\text{T}2}(\cdot; \rho_n^\alpha) + \text{IMSE}$	2 iterations IMSE, 1 iteration with E

Table 1: Strategies implemented in the test cases.

In each test case we choose a tensor product covariance kernel from the Matérn family, see Rasmussen and Williams (2006, Chapter 4), for details on the parametrization.

All computations are carried out in the **R** programming language (R Core Team, 2016), with the packages `DiceKriging` (Roustant et al., 2012) and `DiceDesign` (Franco et al., 2013) for Gaussian modelling, `KrigInv` (Chevalier et al., 2014c) for already existing sampling criterion and `ConservativeEstimates` (Azzimonti and Ginsbourger, 2016) to compute the conservative estimates.

5.1 Benchmark study: Gaussian processes

Let us start with a benchmark study for the different strategies introduced in Section 4 on Gaussian process realizations in two and five dimensions.

The following setup is shared between the two test case. We consider the unit hypercube $\mathbb{X} = [0, 1]^d$, $d = 2, 5$ and we choose a Gaussian process $(Z_x)_{x \in \mathbb{X}} \sim GP(\mathbf{m}, \mathfrak{K})$. The GP has a constant prior mean $\mathbf{m} = 0$ and a tensor product Matérn covariance kernel with parameters fixed as detailed in Table 2. The objective is to obtain a conservative estimate at level $\alpha = 0.95$ for $\Gamma = \{x \in \mathbb{X} : Z_x \geq 1\}$. The measure of reference μ is the Lebesgue measure on \mathbb{X} . Here we test the strategies detailed in Table 1.

We consider an initial design of experiments $\mathbf{X}_{n_{\text{init}}}$, obtained with the function `optimumLHS` from the package `lhs`. The field is simulated at $\mathbf{X}_{n_{\text{init}}}$ and the random seed is saved. The next evaluations are chosen by minimizing each sampling criterion detailed in Table 1. We ensure that each strategy is applied on the same realization by initializing the seed at the same value for sampling $X_{n_{\text{init}}+1}$. Each criterion is run for $n = 30$ iterations, updating the model with $q = 1$ new evaluations at each step. We consider m_{doe} different

Test case	d	covariance parameters	m_{doe}	n_{init}
GP	2	$\nu = 3/2, \theta = [0.2, 0.2]^T, \sigma^2 = 1$	10	3
GP	5	$\nu = 3/2, \theta = \sqrt{\frac{5}{2}}[0.2, 0.2, 0.2, 0.2, 0.2]^T, \sigma^2 = 1$	10	6
Costal	2	$\nu = 5/2, \text{MLE for } \theta, \sigma^2, \sigma_{\text{noise}}^2$	10	10
Nuclear	2	$\nu = 5/2, \text{MLE for } \theta, \sigma^2, \sigma_{\text{noise}}^2$	10	10

Table 2: Test cases parameter choices.

initial design of experiments and, for each design, we replicate the procedure 10 times with different initial values $Z_{\mathbf{x}_{n_{\text{init}}}}$.

5.1.1 Dimension 2

We evaluate the strategies by looking at the type I and type II errors for $Q_{\rho_n^\alpha}$, defined in Section 3, and by computing the measure $\mu(Q_{\rho_n^\alpha})$. For each of these quantities we report the median result over the replications obtained after $n = 30$ evaluations for each initial design.

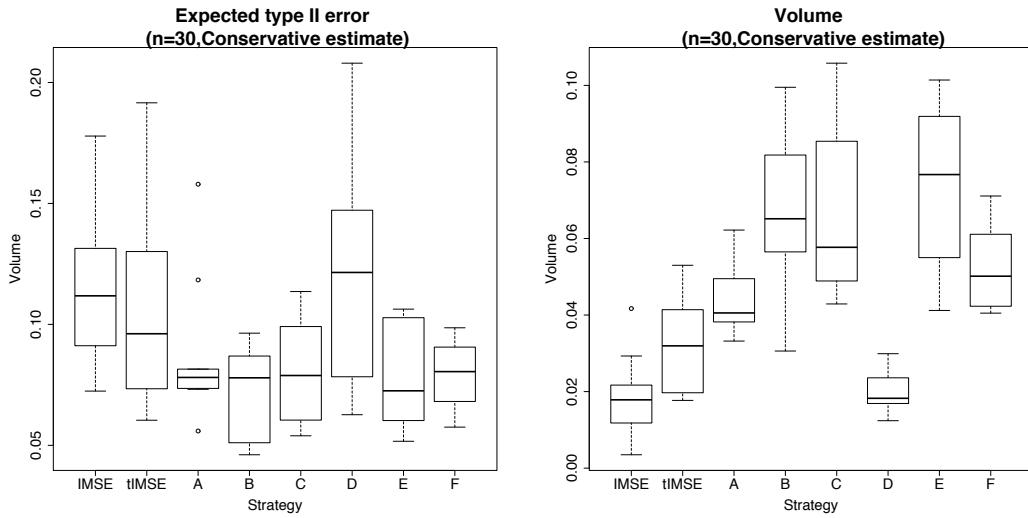
Expected type I error does not vary much among the different strategies as it is controlled by the probabilistic condition imposed on the estimate, as shown in Section 2.

In Figure 1 we show the distribution of expected type II error and the expected volume $\mu(Q_{\rho_n^\alpha})$ after $n = 30$ new evaluations. The strategies A, B, C, E, F all provide better uncertainty reduction for conservative estimates than a standard IMSE strategy or than a tIMSE strategy. In particular strategy E has the lowest type 2 error while at the same time providing an estimate with the largest measure, thus yielding a conservative set which is likely to be included in Γ^* and not too small in volume. All estimates are very conservative: the median ratio between the expected type I error and the estimate's volume is 0.03%, thus much smaller than the upper bound $1 - \alpha = 5\%$ computed in Remark 1. On the other hand the expected type II error is in median 178% bigger than the estimate's volume.

5.1.2 Dimension 5

In Figure 2 we show the distribution of expected type II errors for $Q_{\rho_n^\alpha}$ and its measure $\mu(Q_{\rho_n^\alpha})$ obtained with the different design of experiments, after 30 iterations of each strategy.

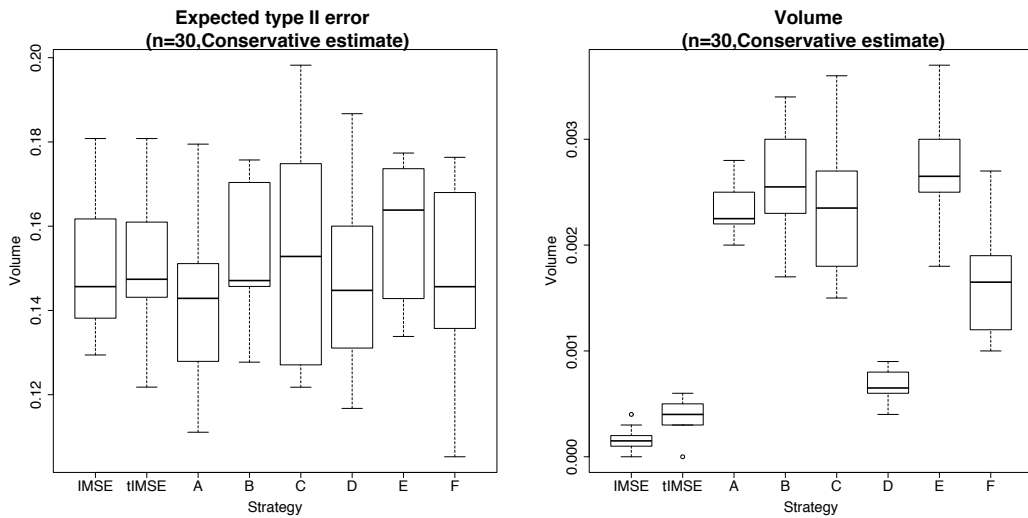
In this test case the differences between the strategies are less clear. The IMSE strategy provides conservative estimates with small measure and with slightly larger type II error. Strategies A, B, C, E provide a good trade off



(a) Median type II error for $Q_{\rho_n^\alpha}$.

(b) Measure $\mu(Q_{\rho_n^\alpha})$.

Figure 1: Median expected type II errors and measure of the estimate across the different designs of experiments after $n = 30$ iterations of each strategy. Test case in dimension 2.



(a) Median type II error for $Q_{\rho_n^\alpha}$.

(b) Measure $\mu(Q_{\rho_n^\alpha})$.

Figure 2: Median expected type II errors and measure of the estimate across different designs of experiments after $n = 30$ iterations of each strategy. Test case in dimension 5.

between small type II error and large measure of the estimate, however they are not clearly better than the other strategies in this case. The estimates provided by all methods are very conservative also in this case. The median ratio over all DoEs and all replications between the expected type I error and volume is 0.33%, which is smaller than the upper bound 5%, as computed in Remark 1. The expected type II error is instead 3 orders of magnitude larger than the estimate’s volume. This indicates that we have only recovered a small portion of the true set Γ^* , however, under the model, this estimate is very conservative.

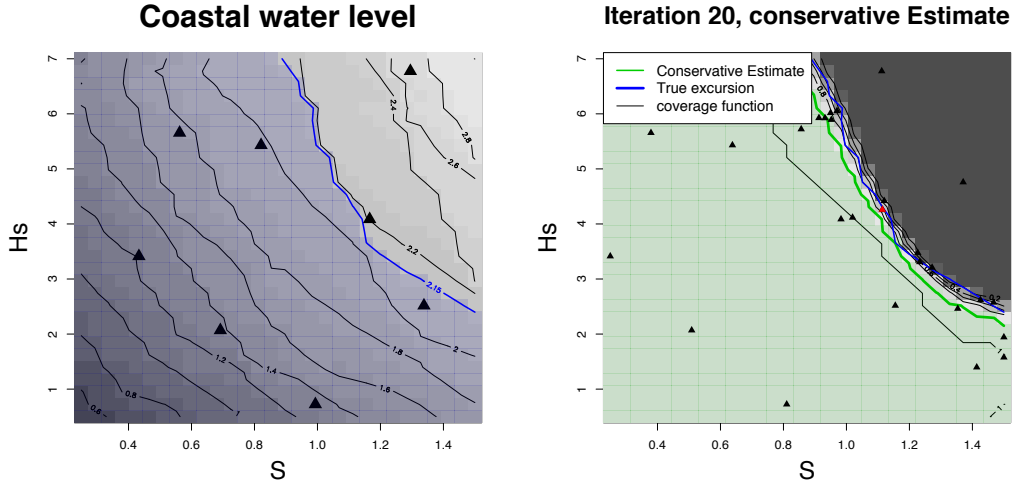
5.2 Coastal flood test case

In this section we present a coastal flood test case introduced in Rohmer and Idier (2012). We focus here on studying the parameters that lead to floods on the coastlines. This study is often conducted with full grid simulations. However, this type of simulations require many hours of computational time and render set estimation problems often infeasible. The use of meta-models, recently revisited in this field (see, e.g., Rohmer and Idier, 2012, and references therein), allows tackling this computational issue.

Here we consider a simplified coastal flood case as described by Rohmer and Idier (2012). The water level at the coast is modelled as a deterministic function $f : \mathbb{X} \subset \mathbb{R}^d \rightarrow \mathbb{R}$, assuming steady offshore conditions, without solving the flood itself inland. The input space $\mathbb{X} = [0.25, 1.50] \times [0.5, 7]$ are the variables storm surge magnitude S and significant wave height H_s . We are interested in recovering the set $\Gamma^* = \{x \in \mathbb{X} : f(x) \leq t\}$, with $t = 2.15$. In order to evaluate the quality of the meta-model, we rely on a grid experiment of 30×30 runs carried out by Rohmer and Idier (2012).

Here we consider a Gaussian process prior $(Z_x)_{x \in \mathbb{X}} \sim GP(\mathbf{m}, \mathfrak{K})$, with constant prior mean function and Matérn covariance kernel with $\nu = 5/2$. We assume that the function evaluations are noisy with zero noise mean and variance σ_{noise}^2 . We select $m_{\text{doe}} = 10$ different initial DoEs, with equal size $n_{\text{init}} = 10$. The initial designs are chosen with a maximin LHS design $\mathbf{X}_{n_{\text{init}}} = \{x_1, \dots, x_{n_{\text{init}}}\} \subset \mathbb{X}$ with the function `optimumLHS` from the package `lhs`. The covariance kernel hyper-parameters and the noise variance are estimated with maximum likelihood. Figure 3a shows the true function f we are aiming at reconstructing, the critical level and one initial design of experiments. We compute conservative set estimates for Γ^* at level $\alpha = 0.95$, as defined in Section 2, with the Lebesgue measure on \mathbb{X} .

We proceed to add 20 evaluations with the strategies detailed in Table 1. The covariance kernel hyper-parameters are re-estimated at each step with maximum likelihood. Figure 3b shows the conservative estimate obtained after



(a) Water level, set of interest (shaded blue, volume=77.56%) and one initial design of experiments (black triangles). (b) Coverage function and conservative estimate (shaded green, $\alpha = 0.95$) after 30 function evaluations (strategy E).

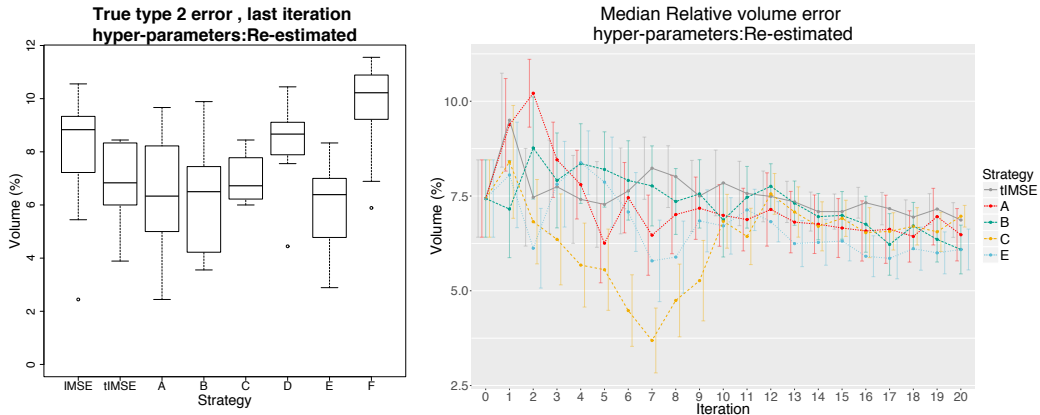
Figure 3: Coastal flood test case, one initial design of experiments.

30 functions evaluations at locations chosen with Strategy *E*.

Figure 4a shows the true type II error at the last iteration of each strategy, after 30 evaluations of the function. The true type II error is computed by comparing the conservative estimate with an estimate of Γ^* obtained from the 30×30 grid experiment. Monte Carlo integration over this grid of evaluations leads to a volume of Γ^* equal to 77.56%.

Strategies *A*, *B*, *E* provide estimates with higher volume and lower type II error in median than IMSE and tIMSE. Strategy *F* did not show any improvement over IMSE or tIMSE. For all strategies the true type I error is zero for almost all initial DoEs, thus indicating that all strategies lead to conservative estimates.

Figure 4b shows the behaviour of relative volume error as a function of the iteration number for Strategies tIMSE, *A*, *B*, *E*. The hyper-parameter re-estimation causes the model to be overconfident at the initial iterations, thus increasing the relative volume error. As the number of evaluations increases the hyper-parameter estimates become more stable and the relative error decreases as conservative estimates are better included in the true set.



(a) True type II error for $CE_{\alpha,30}$ at the last iteration. (b) Relative volume error as a function of iteration number. Strategies tIMSE, A, B, E.

Figure 4: Randomized initial DoEs results. Values of the uncertainties for each strategy with $\alpha = 0.95$. Coastal flood test case.

5.3 Nuclear criticality safety test case

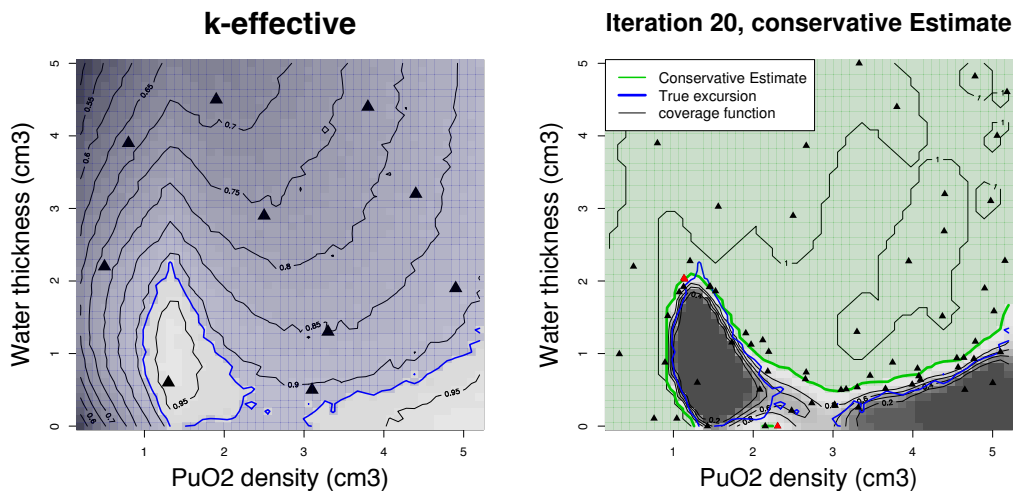
In this section we test the proposed strategies in a reliability engineering test case from the French Institute of nuclear safety (IRSN). The problem at hand concerns a nuclear storage facility and we are interested in estimating the set of parameters that lead to a safe storage of the material. This is closely linked to the production of neutrons. In fact, since neutrons are both the product and the initiator of nuclear reactions, an overproduction could lead to a chain reaction. The safety of a system is usually evaluated with the neutron multiplication factor, here called k -effective or k -eff. We consider $k\text{-eff} : \mathbb{X} \rightarrow [0, 1]$ defined on $\mathbb{X} = [0.2, 5.2] \times [0, 5]$, where the two parameters represent the fissile material density, PuO_2 , and the water thickness, H_2O . We are interested in recovering the set of safe configurations

$$\Gamma^* = \{(\text{PuO}_2, \text{H}_2\text{O}) \in \mathbb{X} : k\text{-eff}(\text{PuO}_2, \text{H}_2\text{O}) \leq 0.92\},$$

where the threshold was chosen at 0.92 for safety reasons.

In general, the evaluation of k -eff at one point requires an expensive computer experiment thus our objective is to provide an estimate for Γ^* from few evaluations of k -eff and to quantify its uncertainty.

We consider the function k -eff as a realization of a Gaussian process and we fix a prior $(Z)_{x \in \mathbb{X}} \sim GP(\mathbf{m}, \mathfrak{K})$ with constant mean function \mathbf{m} and tensor product Matérn covariance kernel with $\nu = 5/2$. We assume that the function evaluations are noisy with zero mean noise and variance σ_{noise}^2 . We consider $m_{\text{doe}} = 10$ different initial DoEs of size $n_0 = 10$, obtained with the



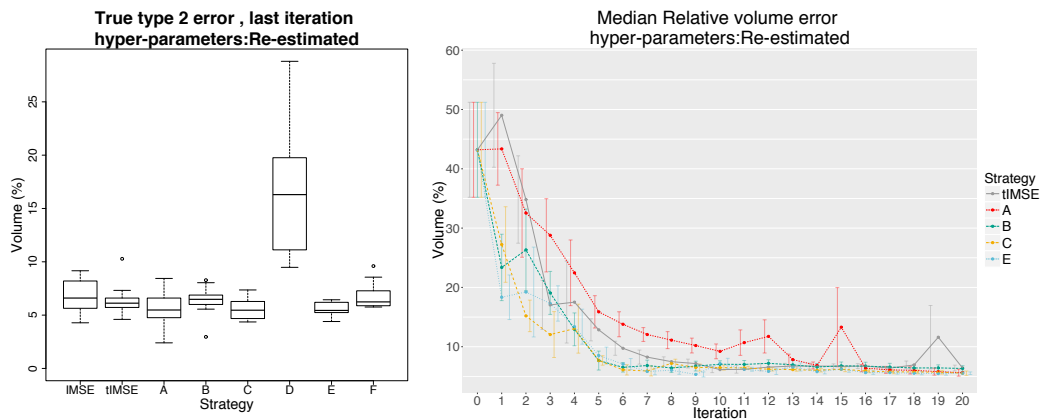
(a) Function k -eff, set of interest Γ^* (shaded blue, volume=87.92%) and one initial design of experiments. (b) Coverage function, conservative estimate ($\alpha = 0.95$, shaded green) after 70 function evaluations (strategy E).

Figure 5: Nuclear criticality safety test case.

function `optimumLHS` from the package `lhs` in **R**. Figure 5a shows the true values of the function, the set Γ^* and one initial design of experiments. Each initial DoE is used to estimate the covariance hyper-parameters and the noise variance with maximum likelihood.

We now test how to adaptively reduce the uncertainty on the estimate with different strategies. Table 1 lists the strategies tested in this section. We run $n = 20$ iteration of each strategy and at each step we select a batch of $q = 3$ new points where to evaluate k -eff. The covariance hyper-parameters are re-estimated at each iteration by adding the 3 new evaluations. The starting model provides a conservative estimate for Γ^* at level $\alpha = 0.95$, with the Lebesgue measure μ on \mathbb{X} . Figure 5b shows the coverage function of the random set Γ obtained after 70 function evaluations at locations selected with Strategy *E* and the corresponding conservative estimate.

Figure 6a shows a comparison of the type II error at the last iteration, i.e. after 70 evaluations of the function, for each initial DoE and each strategy. Strategy *D* as in the previous test cases is the worse performer, thus showing that minimization of type II error works only for conservative quantiles. The strategies *A*, *B*, *E* perform well both in terms of final volume and true type II error. Strategy *F* while better than strategies *D* and IMSE does not show any improvement over its non-hybrid analogous *E*. Strategy *C* also performs well in this test case, as opposed to the previous examples.



(a) True type II error at the final iteration. (b) Relative volume error as a function of the iteration number. Strategies tIMSE, A, B, C, E.

Figure 6: Randomized initial DoEs results. Values of the uncertainties for each strategy with $\alpha = 0.95$. Nuclear criticality safety test case.

Figure 6b shows the relative volume error as a function of the iteration number for strategies tIMSE, A, B, C, E. The relative volume error is computed by comparing the conservative estimate with an estimate of Γ^* obtained from evaluations of k-eff on a grid 50×50 . The volume of Γ^* computed with Monte Carlo integration from this grid of evaluations is 87.92%. All strategies presented show a strong decrease in relative volume error in the first 10 iteration, i.e. until 40 evaluations of k-eff are added. In particular strategies B, C, E show the strongest decline in error in the first 5 iterations. Overall, as in the previous test cases, strategy E, the minimization of the expected type II error, seem to provide the best uncertainty reductions both in terms of relative volume error and in terms of type II error.

6 Conclusion

In this paper we introduced sequential uncertainty reduction strategies for conservative estimates. This type of set estimates proved to be useful in reliability engineering, however they could be of interest in all situations where practitioners aim at controlling the overestimation of the set. The estimator CE, however, is based on a global quantity and an underlying GP model that badly approximates f will not lead to reliable estimates. For a fixed model, this issue might be reduced by increasing the level of confidence. We presented test cases with fixed $\alpha = 0.95$, however testing different levels, e.g. $\alpha = 0.99, 0.995$, and comparing the results is a good practice.

The sequential strategies proposed here provide a way to reduce the uncertainty on conservative estimates by adding new function evaluations. The numerical studies presented showed that adapted strategies provide a better uncertainty reduction than generic strategies. In particular, strategy E , i.e. the criterion $J_n^{T^2}(\cdot; \rho_n^\alpha)$, resulted among the best criteria in terms of Type 2 uncertainty and relative volume error in all test cases. In this work we mainly focused on showing the differences between the strategy with a-posteriori measures of uncertainty. Nonetheless the expected type I and II errors could be used to provide stopping criteria for the sequential strategies. Further studies in this direction are needed to understand the limit behaviour of these quantities as the number of evaluations increases.

The strategies proposed in this work focus on reducing the uncertainties on conservative estimates. This objective does not necessarily lead to better overall models for the function or to good covariance hyper-parameters estimation. The sequential behaviour of hyper-parameters maximum likelihood estimators under SUR strategies needs to be studied in more details.

Acknowledgements

The first author gratefully acknowledges the support of the Swiss National Science Foundation, grant number 146354. The authors wish to thank Jérémy Rohmer (BRGM) and Déborah Idier (BRGM) for the challenging and insightful test case.

References

- Arnaud, A., Bect, J., Couplet, M., Pasanisi, A., and Vazquez, E. (2010). Évaluation d'un risque d'inondation fluviale par planification séquentielle d'expériences. In *42èmes Journées de Statistique de la SFdS*.
- Azzimonti, D., Bect, J., Chevalier, C., and Ginsbourger, D. (2016). Quantifying uncertainties on excursion sets under a Gaussian random field prior. *SIAM/ASA J. Uncertain. Quantif.*, 4(1):850–874.
- Azzimonti, D. and Ginsbourger, D. (2016). Estimating orthant probabilities of high dimensional gaussian vectors with an application to set estimation. *preprint, hal-01289126*.
- Bayarri, M. J., Berger, J. O., Calder, E. S., Dalbey, K., Lunagomez, S., Patra, A. K., Pitman, E. B., Spiller, E. T., and Wolpert, R. L. (2009). Using sta-

- tistical and computer models to quantify volcanic hazards. *Technometrics*, 51(4):402–413.
- Bect, J., Bachoc, F., and Ginsbourger, D. (2016). A supermartingale approach to Gaussian process based sequential design of experiments. *arXiv:1608.01118 [stat.ML]*.
- 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. *Stat. Comput.*, 22 (3):773–793.
- Bichon, B. J., Eldred, M. I. S., Swiler, L. P., Mahadevan, S., and McFarland, J. M. (2008). Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA Journal*, 46(10):2459–2468.
- Bolin, D. and Lindgren, F. (2015). Excursion and contour uncertainty regions for latent Gaussian models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 77(1):85–106.
- Chevalier, C. (2013). *Fast uncertainty reduction strategies relying on Gaussian process models*. PhD thesis, University of Bern.
- Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y. (2014a). Fast kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics*, 56(4):455–465.
- Chevalier, C., Ginsbourger, D., Bect, J., and Molchanov, I. (2013). Estimating and quantifying uncertainties on level sets using the Vorob’ev expectation and deviation with Gaussian process models. In Uciński, D., Atkinson, A., and Patan, C., editors, *mODa 10 – Advances in Model-Oriented Design and Analysis*. Physica-Verlag HD.
- Chevalier, C., Ginsbourger, D., and Emery, X. (2014b). Corrected kriging update formulae for batch-sequential data assimilation. In *Mathematics of Planet Earth*, Lecture Notes in Earth System Sciences, pages 119–122. Springer Berlin Heidelberg.
- Chevalier, C., Picheny, V., and Ginsbourger, D. (2014c). The KrigInv package: An efficient and user-friendly R implementation of kriging-based inversion algorithms. *Comput. Statist. Data Anal.*, 71:1021–1034.
- Chilès, J.-P. and Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty, Second Edition*. Wiley, New York.

- Emery, X. (2009). The kriging update equations and their application to the selection of neighboring data. *Comput. Geosci.*, 13(3):269–280.
- Fleuret, F. and Geman, D. (1999). Graded learning for object detection. In *Proceedings of the workshop on Statistical and Computational Theories of Vision of the IEEE international conference on Computer Vision and Pattern Recognition (CVPR/SCTV)*, volume 2. Citeseer.
- Franco, J., Dupuy, D., Roustant, O., Damblin, G., and Iooss, B. (2013). *DiceDesign: Designs of Computer Experiments*. R package version 1.3.
- French, J. P. and Sain, S. R. (2013). Spatio-temporal exceedance locations and confidence regions. *Ann. Appl. Stat.*, 7(3):1421–1449.
- Ginsbourger, D. and Le Riche, R. (2010). Towards gaussian process-based optimization with finite time horizon. In *mODa 9 – Advances in Model-Oriented Design and Analysis*, pages 89–96. Springer.
- González, J., Osborne, M., and Lawrence, N. D. (2016). GLASSES: Relieving The Myopia Of Bayesian Optimisation. In *19th International Conference on Artificial Intelligence and Statistics*, pages 790–799.
- Jones, D. R., Schonlau, M., and Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *J. Global Optim.*, 13(4):455–492.
- Molchanov, I. (2005). *Theory of Random Sets*. Springer, London.
- Osborne, M. A., Garnett, R., and Roberts, S. J. (2009). Gaussian processes for global optimization. In *3rd international conference on learning and intelligent optimization (LION3)*, pages 1–15.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ranjan, P., Bingham, D., and Michailidis, G. (2008). Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, 50(4):527–541.
- Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian processes for machine learning*. MIT Press.
- Rohmer, J. and Idier, D. (2012). A meta-modelling strategy to identify the critical offshore conditions for coastal flooding. *Natural Hazards and Earth System Sciences*, 12(9):2943–2955.

- Roustant, O., Ginsbourger, D., and Deville, Y. (2012). DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software*, 51(1):1–55.
- Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statist. Sci.*, 4(4):409–435.
- Vazquez, E. and Bect, J. (2009). A sequential bayesian algorithm to estimate a probability of failure. *IFAC Proceedings Volumes*, 42(10):546–550.
- Vorob'ev, O. Y. (1984). Srednemernoje modelirovanie (mean-measure modelling). *Nauka, Moscow*, In Russian.
- Wheeler, M. W., Dunson, D. B., Pandalai, S. P., Baker, B. A., and Herring, A. H. (2014). Mechanistic hierarchical gaussian processes. *J. Amer. Statist. Assoc.*, 109(507):894–904.

A Properties of conservative estimates

In the following, let us denote with (Ω, \mathcal{F}, P) a probability space.

Proof of Proposition 1. Let us consider a measurable set M such that $\mu(M) = \mu(Q_\rho)$. For each $\omega \in \Omega$, we have

$$\begin{aligned} \mu(M\Delta\Gamma(\omega)) - \mu(Q_\rho\Delta\Gamma(\omega)) &= 2\left(\mu(\Gamma(\omega) \cap (Q_\rho \setminus M)) - \mu(\Gamma(\omega) \cap (M \setminus Q_\rho))\right) \\ &\quad + \mu(Q_\rho^C) - \mu(M^C). \end{aligned}$$

By applying the expectation on both sides and by remembering that $\mu(Q_\rho^C) = \mu(M^C)$ we obtain

$$\mathbb{E}[\mu(M\Delta\Gamma) - \mu(Q_\rho\Delta\Gamma)] = \mathbb{E}\left[2\left(\mu(\Gamma \cap (Q_\rho \setminus M)) - \mu(\Gamma \cap (M \setminus Q_\rho))\right)\right] \quad (19)$$

$$= 2 \int_{Q_\rho \setminus M} p_\Gamma(u) d\mu(u) - 2 \int_{M \setminus Q_\rho} p_\Gamma(u) d\mu(u), \quad (20)$$

where the second equality comes from the definition of Q_ρ . Moreover, since $p_\Gamma(x) \geq \rho$ for $x \in Q_\rho \setminus M$ and $p_\Gamma(x) \leq \rho$ for $x \in M \setminus Q_\rho$ we have

$$2 \left[\int_{Q_\rho \setminus M} p_\Gamma(u) d\mu(u) - \int_{M \setminus Q_\rho} p_\Gamma(u) d\mu(u) \right] \geq 2\rho[\mu(Q_\rho \setminus M) - \mu(M \setminus Q_\rho)] \quad (21)$$

$$= 2\rho[\mu(Q_\rho) - \mu(M)] = 0, \quad (22)$$

which shows that Q_ρ verifies Equation (6). \square

Proof of Remark 1. Notice that for all $\omega \in \Omega$ such that $Q_{n,\rho_n^\alpha} \subset \Gamma(\omega)$, we have $G_n^{(1)}(\omega) = 0$. By applying the law of total expectation we obtain

$$\begin{aligned} \mathbb{E}_n[G_n^{(1)}] &= \mathbb{E}_n[G_n^{(1)} \mid Q_{n,\rho_n^\alpha} \subset \Gamma]P(Q_{n,\rho_n^\alpha} \subset \Gamma) \\ &\quad + \mathbb{E}_n[G_n^{(1)} \mid Q_{n,\rho_n^\alpha} \setminus \Gamma \neq \emptyset](1 - P(Q_{n,\rho_n^\alpha} \subset \Gamma)) \\ &\leq 0 + \mathbb{E}_n[G_n^{(1)} \mid Q_{n,\rho_n^\alpha} \setminus \Gamma \neq \emptyset](1 - \alpha) \leq \mu(Q_{n,\rho_n^\alpha})(1 - \alpha). \end{aligned}$$

\square