Supplementary Materials: Adaptive Design of Experiments for Conservative Estimation of Excursion Sets

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

October 25, 2019

1 Complements on conservative estimates

The following result here presented as a corollary of proposition 4, main text, is a well known result Molchanov (2005) for the Vorob'ev expectation.

Corollary 1 (of proposition 4). The Vorob'ev expectation Q_{ρ_V} minimizes the expected distance in measure with Γ among all measurable (deterministic) sets M such that $\mu(M) = \mu(Q_{\rho_V})$. Moreover if $\rho_V \geq \frac{1}{2}$, then the Vorob'ev expectation also minimizes the expected distance in measure with Γ among all measurable sets M that satisfy $\mu(M) = \mathbb{E}[\mu(\Gamma)]$.

Proof of corollary 1. The first statement is a direct application of proposition 4 with $\rho = \rho_V$.

For the second statement, by definition we have that either $\mu(Q_{\rho_V}) = \mathbb{E}[\mu(\Gamma)]$ or $\mu(Q_\rho) < \mathbb{E}[\mu(\Gamma)] \leq \mu(Q_{\rho_V})$ for each $\rho > \rho_V$. In the first case we can directly apply proposition 4. In the second case we can apply the same reasoning as in the proof of proposition 4 however in last step of the proof we need to impose $\rho_V \geq \frac{1}{2}$ for obtaining the result.

□

In general, the Vorob'ev quantile chosen for CE_α is not the set S with the largest measure μ that has the property $P(S \subset \Gamma) \geq \alpha$ as shown in the counterexample below.

*Istituto Dalle Molle di studi sull'Intelligenza Artificiale (IDSIA), Scuola universitaria professionale della Svizzera italiana (SUPSI), Università della Svizzera italiana (USI), Via Cantonale 2c, 6928 Manno, Switzerland

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

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

§Institute of Statistics, University of Neuchâtel, Avenue de Bellevaux 51, 2000 Neuchâtel, Switzerland.

¶Laboratoire des Signaux et Systèmes (UMR CNRS 8506), CentraleSupelec, CNRS, Univ Paris-Sud, Université Paris-Saclay, 91192, Gif-sur-Yvette, France.

‖Institut de Radioprotection et de Sûreté Nucléaire (IRSN), Paris, France.

Example 1. *Consider a discrete set* $\mathbb{X} = \{x_1, x_2, x_3, x_4\}$*, a random field* $(Z_x)_{x\in D}$ *and* $\Gamma = \{x \in \mathbb{X} : Z_x \geq 0\}$*. In this framework we show the existence of a conservative set at level* $\alpha = 0.5$ *larger than the largest Vorob'ev quantile with the same conservative property.*

Assume that for some $\rho_1 \in [0, 1]$

$$P(Q_{\rho_1} \subset \Gamma) = P(Z_{x_1} \geq 0, Z_{x_2} \geq 0) = 1/2,$$

where $Q_{\rho_1} = \{x_1, x_2\}$ *is a Vorob'ev quantile at level* ρ_1*, that is* $P(Z_{x_1} \geq 0), P(Z_{x_2} \geq 0) \geq \rho_1$*.*

Note that in the case where $Z_{x_1} \perp\!\!\!\perp Z_{x_2}$*, the Vorob'ev level is automatically determined as* $\rho_1 = \sqrt{2}/2$ *and if* $Z_{x_1} = Z_{x_2}$ *a.s., then* $\rho_1 = 1/2$*. Let us assume here that* $Z_{x_1} \neq Z_{x_2}$*, which implies* $\rho_1 \in (1/2, \sqrt{2}/2)$*. Let us further denote with* Ω_1 *the subset of* Ω *such that for all* $\omega \in \Omega_1$ $\min(Z_{x_1}(\omega), Z_{x_2}(\omega)) \geq 0$ *and define* $\Omega_2 = \Omega \setminus \Omega_1$*.*

We further fix Z_{x_3} *as the random variable*

$$Z_{x_3}(\omega) = \begin{cases} 1 & \text{if } \omega \in \Omega_1 \\ -1 & \text{if } \omega \in \Omega_2. \end{cases}$$

Then $P(Z_{x_3} \geq 0) = P(\min(Z_{x_1}, Z_{x_2}) \geq 0) = P(Z_{x_1} \geq 0, Z_{x_2} \geq 0) = 1/2$*. Moreover* $P(\min(Z_{x_1}, Z_{x_2}, Z_{x_3}) \geq 0) = 1/2$*, i.e.* $\{x_1, x_2, x_3\}$ *has the conservative property at level* $\alpha = 0.5$*.*

Consider $\Omega_3 \subset \Omega_1$ *with* $P(\Omega_3) = 1/3$ *and* $\Omega_4 \subset \Omega_2$ *with* $P(\Omega_4) = 1/3$*. Define*

$$Z_{x_4}(\omega) = \begin{cases} 1 & \text{if } \omega \in \Omega_3 \cup \Omega_4 \\ -1 & \text{otherwise.} \end{cases}$$

We now have that $P(\min(Z_{x_1}, Z_{x_2}, Z_{x_3}, Z_{x_4}) \geq 0) = 1/3 < 1/2$ *and* $P(Z_{x_4} \geq 0) = 1/3 + 1/3 > 1/2$*. Under this construction the Vorob'ev quantiles are* $Q_{\rho_1} = \{x_1, x_2\}, Q_{\rho_2} = \{x_1, x_2, x_4\}$ *and* $Q_{0.5} = D$*. The set* $\{x_1, x_2, x_3\}$ *is therefore the conservative set at level* $\alpha = 0.5$*, as it is the largest subset of* D *with the conservative property, however it is not a Vorob'ev quantile.*

2 Sequential conservative excursion set estimation: procedure overview

Consider a function $f : \mathbb{X} \to \mathbb{R}$, we are interested in estimating

$$\Gamma(f) = \{x \in \mathbb{X} : f(x) \geq t\}, \quad t \in \mathbb{R}.$$

from few evaluations. We consider a prior Gaussian process $(Z_x)_{x\in\mathbb{X}}$ with prior mean $\mathfrak{m}$ and covariance kernel $\mathfrak{K}$. The estimation procedure often starts with a small initial design $\mathbf{X}_n$, $n \geq 1$, where n is often chosen as a function of the input space dimension. As a rule of thumbs, the initial number of evaluations is often $n = 10d$. In our framework, often, the initial design is chosen as space filling, such as a Latin hypercube sample (LHS) design or points from a low discrepancy sequence such as the Halton and the Sobol' sequence.

In algorithm 1 we summarize the main steps for computing conservative estimates and evaluating their uncertainties.

Algorithm 1 Sequential conservative excursion set estimation.

Require: N_{tot} maximum number of evaluations, n size of initial design, q size of batches, function f, threshold t, criterion J, uncertainty function H

1: select initial DoE $\mathbf{X}_n$, e.g., with space filling design
2: evaluate the function f at $\mathbf{X}_n$
3: compute the posterior model $Z \mid \mathcal{A}_n$ and the estimate Q_{n,ρ_n^α}
4: $i = n_0$
5: **while** i less than N_{tot} **do**
6: select $\widehat{\mathbf{x}^{(q)}}$ by minimizing $J_i(\mathbf{x}^{(q)})$
7: evaluate the function f at $\widehat{\mathbf{x}^{(q)}}$
8: update the posterior model $Z \mid \mathcal{A}_{i+q}$
9: compute the conservative estimate Q_{i+q,ρ_i^α}
10: evaluate the uncertainty function H_{i+q} on Q_{i+q,ρ_i^α}
11: $i = i + q$
12: **end while**
13: optional post-processing on $Q_{N_{tot},\rho_{N_{tot}}^\alpha}$
14: **return** $Q_{N_{tot},\rho_{N_{tot}}^\alpha}$ for $\Gamma(f)$ and the uncertainty value H_N

3 Uncertainty MEAS and related SUR strategy

The measure of a conservative estimate gives a good indication of the uncertainty on the current estimate. For α close to 1, Q_{n,ρ_n^α} is constrained to be inside Γ with high probability, therefore the estimate is often smaller (in measure) than Γ. This allows us to define the following additional uncertainty: the expected difference between the measure of Γ and the measure of Q_{n,ρ_n^α}.

Definition 1 (Uncertainty MEAS)**.** *We denote the uncertainty related to the measure μ with $H^{\text{MEAS}}_{n,\rho_n^\alpha}$, defined as*

$$H^{\text{MEAS}}_{n,\rho_n^\alpha} := \mathbb{E}_n[\mu(\Gamma) - \mu(Q_{n,\rho_n^\alpha})] \tag{1}$$

This quantity is a reasonable uncertainty function only for conservative estimates. In this case, in fact, this quantity is equal to $\mathbb{E}_n[G_n^{(2)}(\rho_n^\alpha) - G_n^{(1)}(\rho_n^\alpha)]$ and, if the estimate is completely included in Γ, then it is the Type II uncertainty.

As for the other criteria defined in section 3, main text, we can use the uncertainty H_n^{MEAS} to define a SUR criterion. Let us denote with J_n^{MEAS} the following function

$$J_n^{\text{MEAS}}(\mathbf{x}^{(q)}; \rho_{n+q}^\alpha) = \mathbb{E}_{n,\mathbf{x}^{(q)}}\left[\mu(\Gamma) - \mu(Q_{n+q,\rho_{n+q}^\alpha})\right]. \tag{2}$$

Since we are interested in minimizing this criterion and $\mathbb{E}_n[\mu(\Gamma)]$ is independent from $\mathbf{x}^{(q)}$, we consider the equivalent function to maximize

$$\widetilde{J_n}^{\text{MEAS}}(\mathbf{x}^{(q)}; \rho_{n+q}^\alpha) = \mathbb{E}_{n,\mathbf{x}^{(q)}}\left[\mu(Q_{n+q,\rho_{n+q}^\alpha})\right].$$

Table 1: MC function evaluation scenarios, total cost $O(n_{MC}kq)$ fixed.

q	τ^2	n_{MC}	k	n
1	0.05	20	80	80
8	0.25	4	50	400
16	0.5	2	50	800

Note minimizing $\widetilde{J_n}^{\text{MEAS}}$ selects points that are meant to increase the measure of the estimate and it is only reasonable for conservative estimates where the conservative condition leads to $Q_{n+q,\rho^\alpha_{n+q}}$ with finite measure in expectation.

In the particular case where $T = (-\infty, t]$ this criterion has a closed-form formula.

Proposition 1 (Measure criterion)**.** *The criterion J_n^{MEAS} can be expanded in closed-form as*

$$\begin{aligned}\widetilde{J_n}^{\text{MEAS}}(\mathbf{x}^{(q)}; \rho_n^\alpha) &= \mathbb{E}_{n,\mathbf{x}^{(q)}}\left[\mu(Q_{n+q,\rho_n^\alpha})\right] \\ &= \int_{\mathbb{X}} \Phi\left(\frac{a_{n+q}(u) - \Phi^{-1}(\rho_n^\alpha)}{\sqrt{\gamma_{n+q}(u)}}\right) d\mu(u). \end{aligned} \tag{3}$$

Proof of proposition 1. The indicator function of the set Q_{n+q,ρ_n^α} can be written as $\mathbb{1}_{p_{n+q}(x)\geq\rho_n^\alpha}$. By Tonelli's theorem we exchange the expectation with the integral over $\mathbb{X}$ and we obtain

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

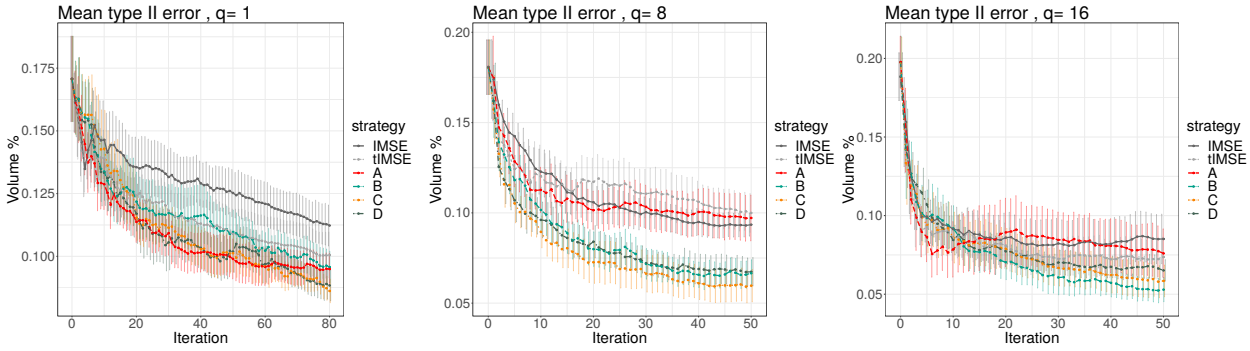
By substituting the expression in equation equation (17) we obtain

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

□

4 Batch-sequential strategies in high noise scenarios

In this section we consider the synthetic test case introduced in section 4, main text, and we run an additional benchmark study with higher noise levels. Table 1 describes the allocation of resources in this benchmark. Note that the noise variance τ^2 is 500 times larger here than the benchmark in the main text.



(a) $q = 1$. (b) $q = 8$. (c) $q = 16$.

Figure 1: Expected Type II error in different batch sequential scenarios.

As in section 4, main text, we consider the unknown function f as a realization of $(\xi_x)_{x\in\mathbb{X}} \sim GP(\mathfrak{m}, \mathfrak{K})$ with constant mean function $\mathfrak{m}$ and Matérn covariance kernel $\mathfrak{K}$ with smoothness parameter $\nu = 3/2$, variance $\sigma^2 = 1$ and lengthscales $\theta_i = 0.2$, $i = 1, 2$. Also in this case the set to estimate is $\Gamma(f) = \{x \in [0,1]^2 : f(x) \geq 1\}$, an excursion above $t = 1$ and μ is the usual volume on $[0,1]^2$. For each scenario we consider an initial DoE of size $n_{\text{init}} = 3$ and we select the next function evaluation with the strategies listed in table 3, main text, with the additional strategy D, where the criterion $\widetilde{J}_n^{\text{MEAS}}$ introduced above is maximized. We run each strategy for the number of iteration specified in table 1. We consider $m_{\text{doe}} = 10$ different initial DoE and, for each design, we replicate the procedure 10 times with different values for $\xi_{\mathbf{X}_{\text{init}}}$.

Figure 1 shows the expected type II error for each strategy in the two batch sequential scenarios. Note that while the convergence of the parallel scenario ($q = 8$) is still faster than the sequential one, here the difference is much less important than in the less noisy example. The scenario $q = 16$ shows a slightly faster convergence for all strategies, however the difference with $q = 8$ is very small. A smarter online allocation strategy such as the one outlined in Picheny et al. (2013) could further improve the performances of batch-sequential strategies.

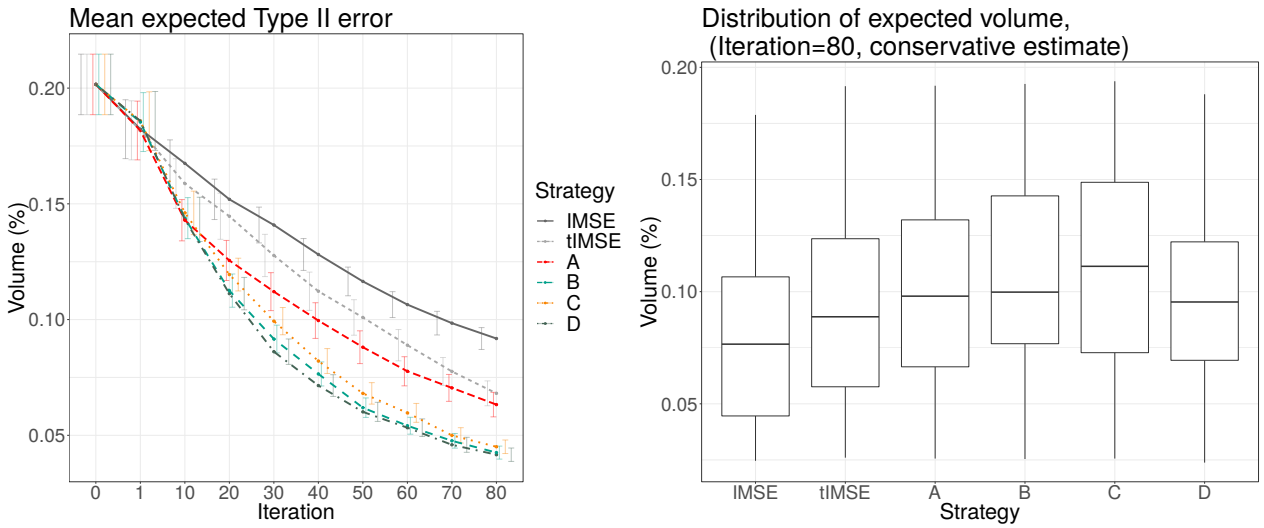
5 Noise free numerical benchmarks

In this section we develop a benchmark study with Gaussian process realizations to study the different behavior of the proposed strategies. We consider two cases with the following shared setup. The input space is the unit hypercube $\mathbb{X} = [0,1]^d$, for $d = 2, 5$ and $(\xi_x)_{x\in\mathbb{X}} \sim GP(\mathfrak{m}, \mathfrak{K})$ with constant prior mean $\mathfrak{m} \equiv 0$ and tensor product Matérn covariance function with known hyper-parameters fixed as in table 2. The noise variance here is constant and equal to zero. The objective is a conservative estimate at level $\alpha = 0.95$ for $\Gamma = \{x \in \mathbb{X} : \xi_x \geq 1\}$ and μ is the usual volume. We test the strategies listed in table 3, main text, and the additional strategy D which maximizes the MEAS criterion.

We consider an initial design of experiments $\mathbf{X}_{n_{\text{init}}}$, obtained with the function `optimumLHS` from the package `lhs` and we simulate the field at $\mathbf{X}_{n_{\text{init}}}$. The size n_{init} (see table 2) is cho-

Table 2: Test cases parameter choices.

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	$\nu = 3/2$, $\theta = [0.2, 0.2, 0.2, 0.2, 0.2]^T$, $\sigma^2 = 1$	10	6



(a) Mean type II error for Q_{n,ρ_n^α} across different initial DoE, $n = 80$ iterations.

(b) Volume $\mu(Q_{n,\rho_n^\alpha})$ across different initial DoE, after $n = 80$ iterations.

Figure 2: Gaussian process realizations test case in dimension 2.

sen small to highlight the differences between the sequential strategies. We select the next evaluations by minimizing each sampling criterion detailed in table 3, main text, and by maximizing $\widetilde{J}_n^{\text{MEAS}}(\cdot; \rho_n^\alpha)$. Each strategy is run for $n = 80$ ($n = 120$ if $d = 5$) iterations, updating the model with $q = 1$ new evaluations at each step. We consider m_{doe} different initial design of experiments and, for each design, we replicate the procedure 10 times with different initial values $\xi_{\mathbf{x}_{n_{\text{init}}}}$.

We evaluate the strategies by looking at the type I and type II errors for Q_{n,ρ_n^α}, defined in section 3.1, main text, and by computing the measure $\mu(Q_{n,\rho_n^\alpha})$. Since the estimate Q_{n,ρ_n^α} has a guaranteed low type I error, a large measure is an indicator of a non trivial conservative estimate. We report mean and median result for each initial design. Expected type I error does not vary much among the different strategies as it is controlled by the condition defining conservative estimate, as shown in section 3.1, main text.

5.1 Dimension 2 GP realizations

Figure 2a shows the expected type II error at selected iteration numbers averaged across different initial DoE. This quantity decreases for all strategies, however strategy B and C outperform the others. Figure 2b shows the values of expected volume $\mathbb{E}_n[\mu(Q_{n,\rho_n^\alpha})]$ obtained after $n = 80$ new evaluations, across different initial DoEs.

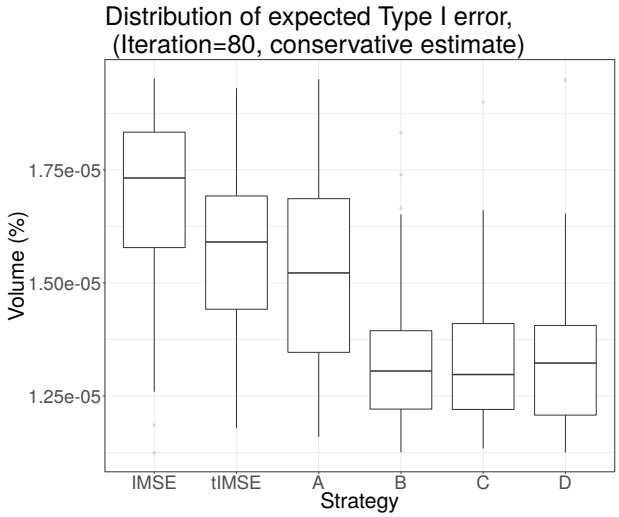


Figure 3: Median type I error for Q_{n,ρ_n^α}, Gaussian processes test case, $d = 2$.

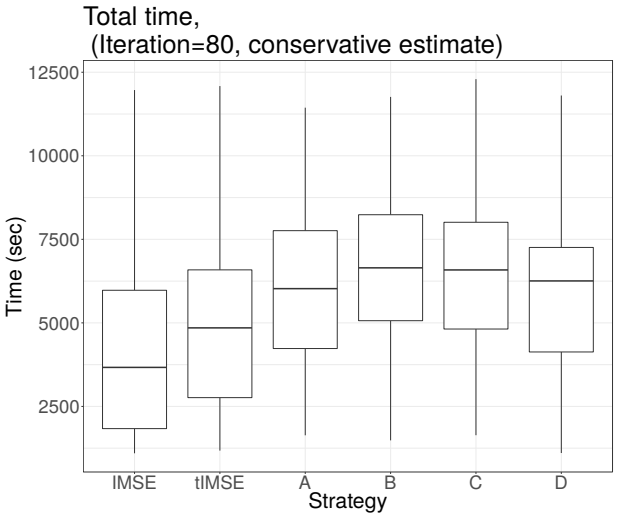


Figure 4: Total time to compute Q_{n,ρ_n^α}, Gaussian processes test case, $d = 2$.

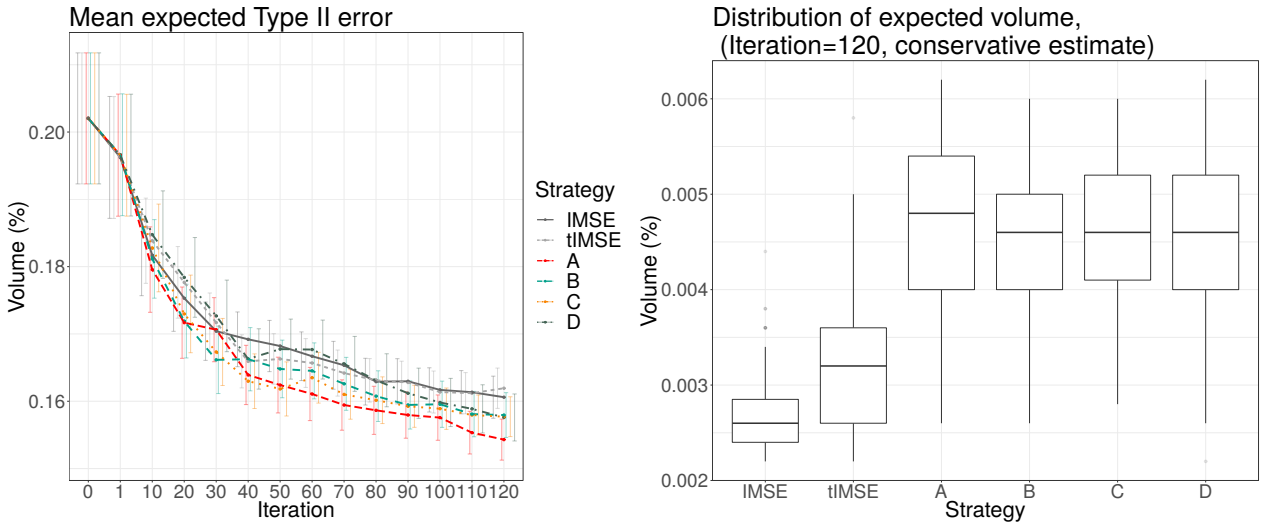
The strategies A, B, C, D all provide better uncertainty reduction for conservative estimates than a standard IMSE strategy or than a tIMSE strategy. In particular strategy C has the second lowest mean type 2 error while at the same time providing an estimate with the largest measure, thus yielding a conservative set likely to be included in $\Gamma(f)$ and, at the same time, not trivial. All estimates, however, are very conservative: the final median ratio between the expected type I error and the estimate's volume is 0.016%, much smaller than the upper bound $1 - \alpha = 5\%$ computed in proposition 1, main text. On the other hand, the median ratio between the expected type II error and the volume at the last iteration is between 31% (C) and 143% (IMSE).

Figure 3 shows the type I error for each strategy in table 3, main text, plus strategy D, i.e., maximization of $\widetilde{J_n^{\text{MEAS}}}$. Strategies B and C show a lower type I error with respect to the other strategies, however the all strategies present very low type I error compared to the total expected measure of the set.

Figure 4 shows the total time required to evaluate the criteria and to compute at each step the conservative estimate. The computational time is mainly driven by the size of the conservative estimate. In fact, for conservative estimates with large measure, the othant probabilities involved in its computation are higher dimensional.

5.2 Dimension 5 GP realizations

Figures 5a and 5b show the mean expected type II error over selected iterations and the expected measure $\mathbb{E}_n[\mu(Q_{n,\rho_n^\alpha})]$ after 120 iterations of each strategy. Strategies A, B, C, D provide better uncertainty reduction for conservative estimates than IMSE or tIMSE. Strategies A and C provide a faster reduction of the type II error and a smaller final mean value than the others with strategy A obtaining a slightly higher median value for the expected measure at iteration 120. Also in this case, even if the iteration number is higher, the final estimates provided by all methods are very conservative. Over all DoEs and replications,



(a) Mean type II error for Q_{n,ρ_n^α} across different initial DoE, $n = 120$ iterations.

(b) Volume $\mu(Q_{n,\rho_n^\alpha})$ across different initial DoE, after $n = 120$ iterations.

Figure 5: Gaussian process realizations test case in dimension 5.

the median ratio between the expected type I error and the volume of Q_{n,ρ_n^α} is 0.02%, much smaller than the upper bound 5%. 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 $\Gamma(f)$ and this estimate is very conservative.

Figure 6 shows the type I error and figure 7 shows the total time required to evaluate the criteria and to compute at each step the conservative estimate. The conservative strategies also in this case strategy present smaller type I error, however also in this case the type I error is much smaller than the expected measure.

Figure 7 shows the total time required to evaluate the criteria and to compute at each step the conservative estimate.

5.3 Model-free comparison of strategies

The metrics presented in the previous sections are based on the GP model. In this section we compare the strategies with a simpler metric independent from the underlying model.

We consider the number of evaluation points that are selected inside and outside the excursion set. At each iteration i, this quantity is computed as $\frac{\#\{j:z_j \geq t,\ j=1,\ldots,n_i\}}{n_i}$, where n_i is the total number of points at iteration i and $z_1, \ldots, z_{n_i}$ are the evaluations. Figure 8 shows the proportion of points inside the excursion set at each iteration for the two GP test cases. Strategy IMSE is a space filling strategy therefore the proportion of points inside the excursion reflects the volume of excursion. Strategies A and tIMSE are adapted to the problem of estimating an excursion set, however they are not adapted for conservative estimation, as such they tend to select points around the boundary of Γ and not inside. Strategies B, C and D instead select more points inside the excursion leading to a good trade-off between a good global approximation of the set and a good approximation of the

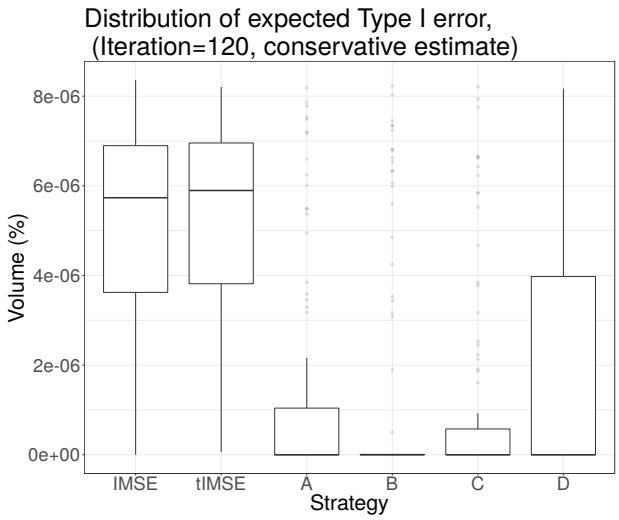


Figure 6: Median type I error for Q_{n,ρ_n^α}, Gaussian processes test case, $d = 5$.

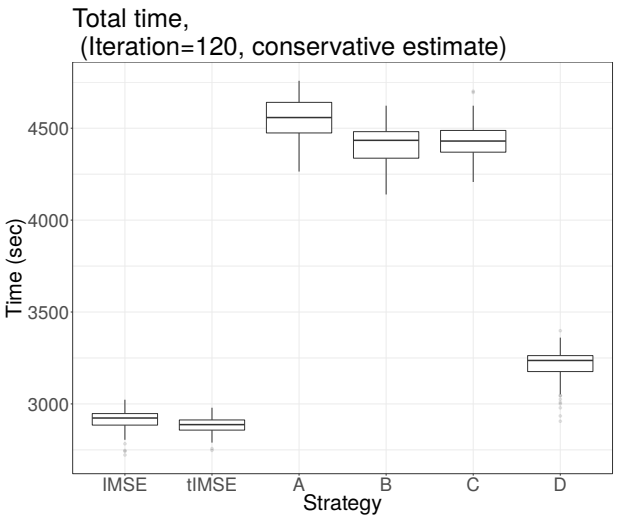
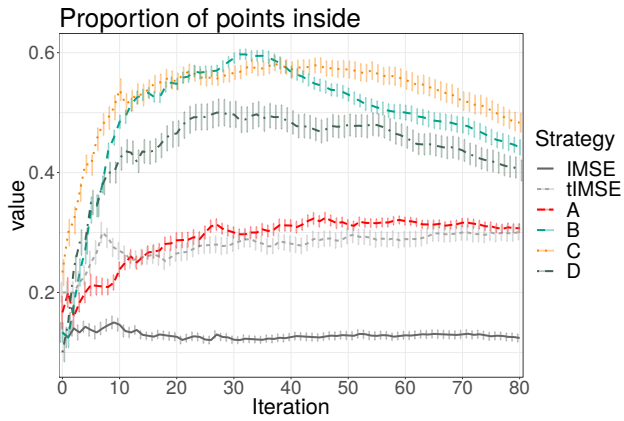
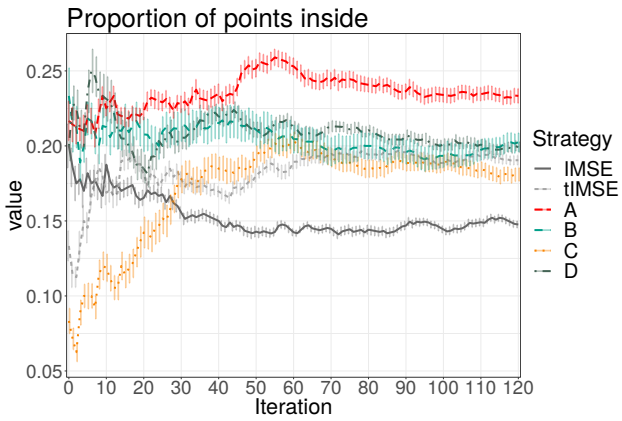


Figure 7: Total time to compute Q_{n,ρ_n^α}, Gaussian processes test case, $d = 5$.

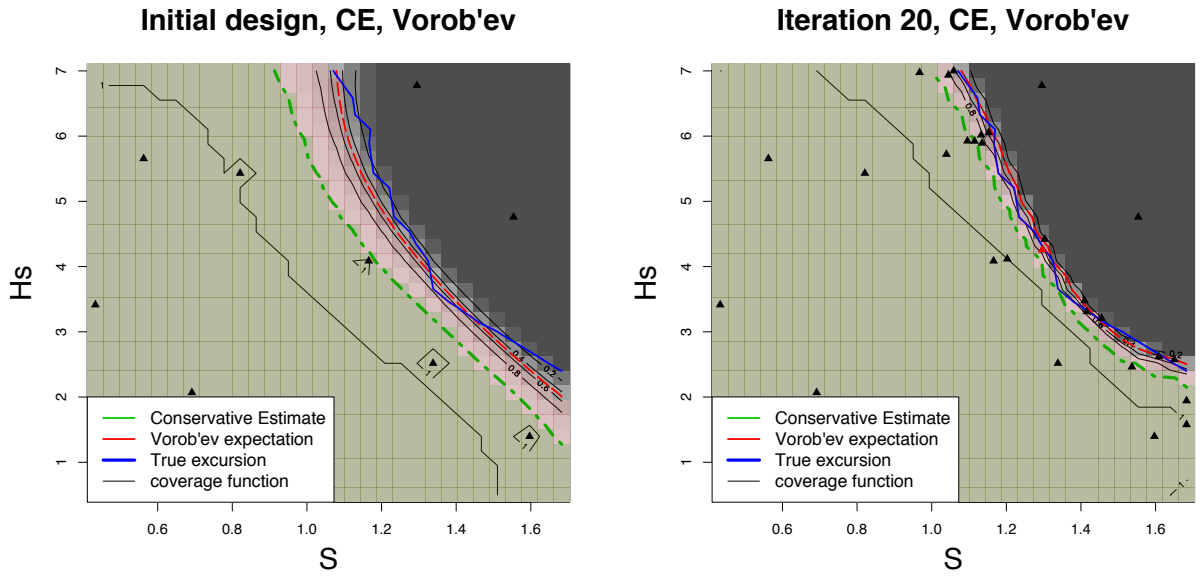


(a) Test case $d = 2$.



(b) Test case $d = 5$.

Figure 8: GP realizations. Average proportion of points inside the excursion region.



(a) One initial DoE (black triangles, $n = 10$), initial coverage function, $\mathrm{CE}_{\alpha,10}$ (shaded green) and Q_{10,ρ_V}, shaded red.

(b) Coverage function, $\mathrm{CE}_{\alpha,30}$ (shaded green, $\alpha = 0.95$) and Q_{30,ρ_V}, final 20 evaluations chosen with strategy C.

Figure 9: Coastal flood test case. Set of interest delimited by blue line, $\mu(\Gamma(f)) = 77.56\%$.

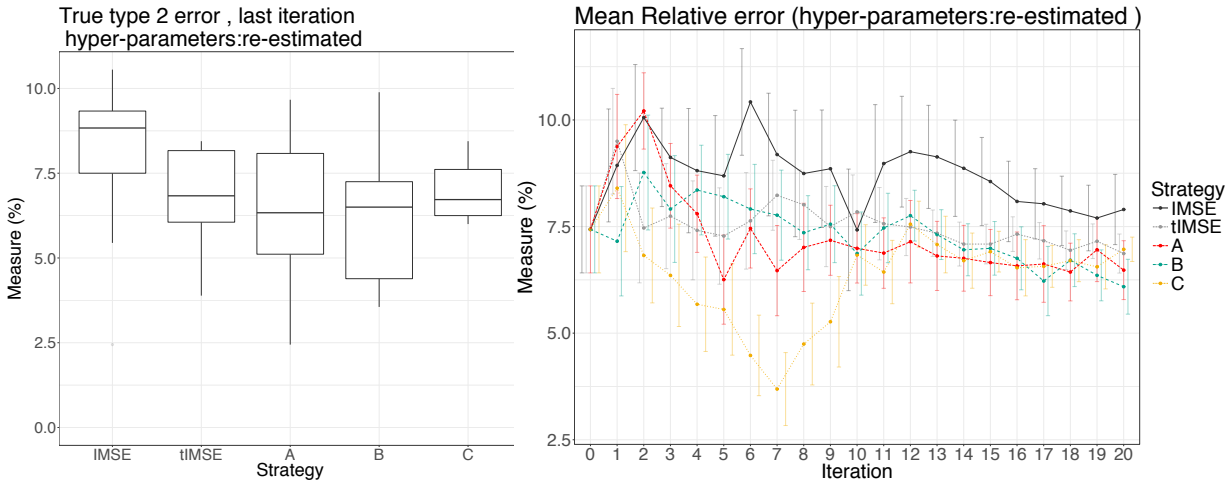
boundary. These observations are reflected in two dimensions, figure 8a, by the proportion of points inside the excursion set. In the five dimensional test case, figure 8b, the proportion of points inside the excursion set is similar across all strategies, except for the IMSE strategy which tends to have a smaller proportion.

5.4 Coastal flood test case

In this section we show how sequential conservative estimates can locate a region of excursion in the coastal flood test case introduced in Rohmer and Idier (2012).

Here we consider the simplified coastal flood case described in Rohmer and Idier (2012). The water level at the coast is modeled as a deterministic function $f : \mathbb{X} \subset \mathbb{R}^d \to \mathbb{R}$, by assuming steady offshore conditions and 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 $\Gamma(f) = \{x \in \mathbb{X} : f(x) \leq t\}$, with $t = 2.15$, shown in figure 9a. In order to evaluate the quality of the meta-model, we rely on the grid experiment of 30×30 runs carried out by Rohmer and Idier (2012).

Here we consider a Gaussian process prior $(Z_x)_{x\in\mathbb{X}} \sim GP(\mathfrak{m}, \mathfrak{K})$, with constant prior mean function and Matérn covariance kernel with $\nu = 5/2$ with MLE hyper-parameters. We assume that the function evaluations are noisy with homogeneous variance σ^2_{noise} estimated from the data. We select $m_{\text{doe}} = 10$ different initial DoEs with a maximin LHS design (function `optimumLHS` from the package `lhs`), with equal size $n_{\text{init}} = 10$. We compute conservative set estimates for $\Gamma(f)$ at level $\alpha = 0.95$, as defined in section 2.1, main text,



(a) True type II error for $\text{CE}_{\alpha,30}$ at the last iteration.

(b) Relative volume error as a function of iteration number. Strategies tIMSE, A, B, C.

Figure 10: Randomized initial DoEs results. Values of the uncertainty functions for each strategy with $\alpha = 0.95$. Coastal flood test case.

with the Lebesgue measure on $\mathbb{X}$.

We proceed to add 20 evaluations with the strategies detailed in table 3, main text. The covariance hyper-parameters are re-estimated at each step with maximum likelihood. Figure 9b shows the conservative estimate obtained after 30 functions evaluations at locations chosen with Strategy C.

Figure 10a 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 $\Gamma(f)$ obtained from the 30×30 grid experiment. Monte Carlo integration over this grid of evaluations leads to a volume of $\Gamma(f)$ equal to 77.56%.

At the last iteration, strategies A, B, C provide estimates with higher volume and lower type II error in median than IMSE and tIMSE. For example, the median type II error for Strategy C is 38% smaller than the IMSE type II error. 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 10b shows the behavior of relative volume error as a function of the iteration number for Strategies tIMSE, A, B, C. 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.

5.5 Hyper-parameter estimation

In this section we explore the behavior of the strategies under different scenarios for the covariance hyper-parameters:

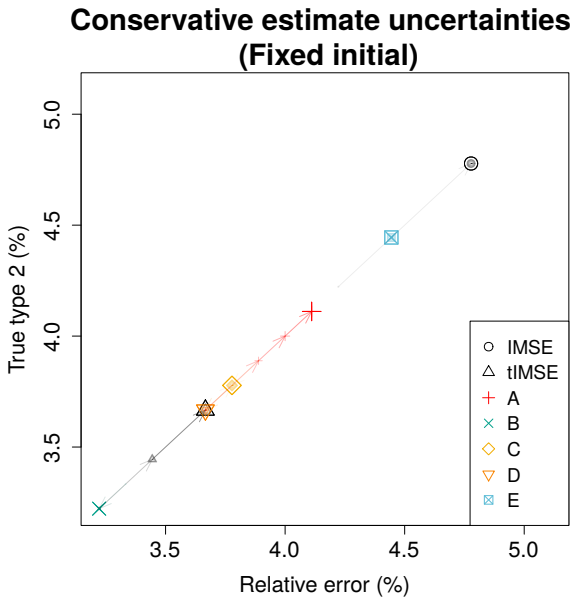


Figure 11: Type II error versus relative volume error for Q_{n,ρ_n^α} for the last 5 iterations, no re-estimation of the covariance parameters.

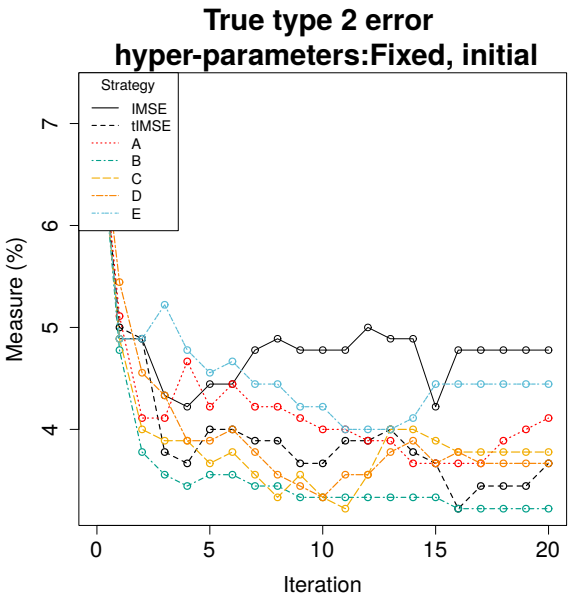


Figure 12: Type II error of Q_{n,ρ_n^α} at each n computed with respect to the true set, no re-estimation of the covariance parameters.

1. *fixed initial hyper-parameters (FI)*: the covariance parameters are fixed throughout the sequential procedure as the maximum likelihood estimates obtained from the initial evaluations $\mathbf{f}_{10}$;

2. *re-estimated hyper-parameters (RE)*: at each step the hyper-parameter estimates are updated with the new evaluations of the function;

We consider the same experimental setup of section 5.4, we fix only one initial DoE of $n = 10$ points chosen with the function `lhsDesign` from the package `DiceDesign` Franco et al. (2013). We run $n = 20$ iterations of each strategy in table 3, main text, where at each iteration we select one evaluation of f. Additionally here we also run strategy D, introduced in section 3, and an hybrid strategy. This strategy, denoted with E, selects points by minimizing alternatively $J_n^{\mathrm{T}2}(\cdot; \rho_n^\alpha)$ for 1 iteration and then IMSE for 2 iterations.

Let us start with the case where we estimate the hyper-parameters only using the evaluations of f at $\mathbf{X}_{10}$. Figure 11 and figure 12 show the type II error and the relative total volume error. These errors are computed comparing the conservative estimate with the true set as in section 5.4. True type I error, not shown, is equal to zero for each strategy at each iteration. Type II error decreases as a function of the evaluations number for all strategies, in particular strategies B, tIMSE, D and C provide a good uncertainty reduction. In particular strategies B, D and C provide a larger uncertainty reduction in the first 10 iterations compared to the other strategies.

In practice, the covariance parameters are often re-estimated after a new evaluation is added to the model. This technique should improve the model, however better conservative estimates are obtained only if the hyper-parameter estimation is stable and reliable. Con-

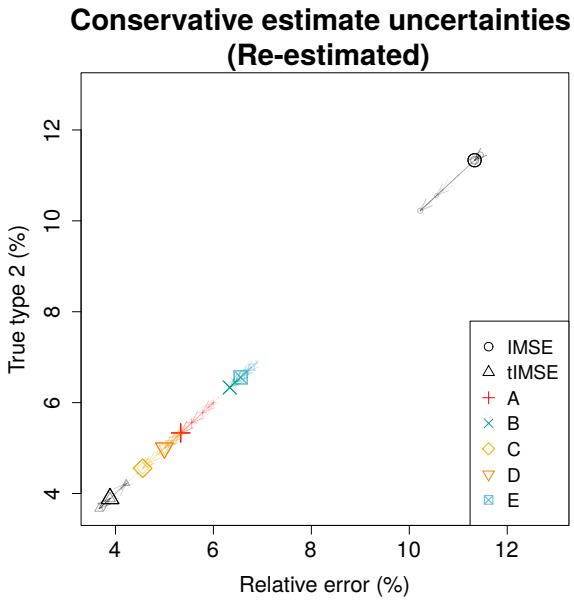


Figure 13: True type II error versus relative volume error for Q_{n,ρ_n^α} at each n, re-estimation of the covariance parameters at each step.

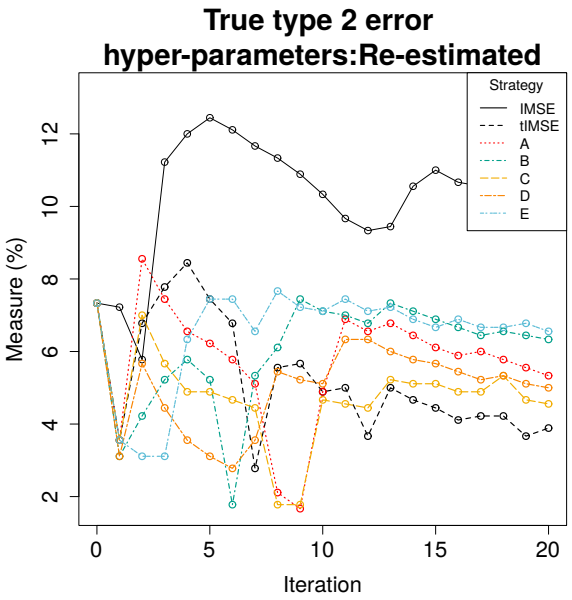


Figure 14: True type II error for Q_{n,ρ_n^α} computed with respect to the true set at each n, re-estimation of the covariance parameters at each step.

servative estimates are in fact based on the coverage probability function and in particular on high quantiles of this function.

Figure 13 and figure 14 show Type II and relative volume error computed comparing the conservative estimate to the true set in the case where covariance parameters are re-estimated at each step. During the first 10 iterations, all strategies except IMSE have small (less than 1%) type I error, not shown, which becomes equal to zero for all strategies after iteration 10.

The strategy IMSE is still the worst performer both in terms of true type II error and of relative volume error, however the remaining strategies do not show big differences. The tIMSE show the best behavior closely followed by Strategy C, D, A. The differences between the final estimated set obtained with these four strategies are small and they are mainly due to a difference in the hyper-parameter estimation. The tIMSE strategy produces more stable hyper-parameter estimators than Strategy C, where the range parameters decrease in the last steps. This change leads to smaller more conservative set estimates.

If the covariance hyper-parameters are kept fixed, figure 12 shows that the true type II error tends to stabilize because the conservative estimate is the best according to the current model. On the other hand parameter re-estimation leads to a more unstable type II error which also indicates that the underlying model is adapting to the new observations.

Hyper-parameters re-estimation at each steps might lead to instabilities also in the maximum likelihood estimators themselves. The MLEs for hyper-parameters in this test case are quite stable, however further studies are required to better understand the behavior of such estimators in a sequential framework.

References

Franco, J., Dupuy, D., Roustant, O., Damblin, G., and Iooss, B. (2013). *DiceDesign: Designs of Computer Experiments*. R package version 1.3.

Molchanov, I. (2005). *Theory of Random Sets*. Springer, London.

Picheny, V., Ginsbourger, D., Richet, Y., and Caplin, G. (2013). Quantile-based optimization of noisy computer experiments with tunable precision. *Technometrics*, 55(1):2–13.

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.