# Sampling criteria for excursion set estimation on multi-output models

C. Duhamel [†,1], C. Prieur [§,1], C. Helbert [§,2], M. Munoz Zuniga [§,3], D. Sinoquet [§,3]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: Oct. 2020 – Jun. 2024*

[1] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK
clement.duhamel@inria.fr ; clementine.prieur@univ-grenoble-alpes.fr

[2] Univ. Lyon, Ecole Centrale de Lyon, CNRS UMR 5208, Institut Camille Jordan,
celine.helbert@ec-lyon.fr

[3] IFP Energies Nouvelles, Rueil-Malmaison
delphine.sinoquet@ifpen.fr ; miguel.munoz-zuniga@ifpen.fr

## Abstract

Many industrial issues are related to excursion set estimation problems, formulated as the estimation of a set of feasible black box model input values, that is the set of model input values satisfying a constraint on a model output, for example to remain below a fixed threshold (see for example [2]). An effective way to solve this problem is to model the costly black-box function of interest as a realization of a Gaussian process (GP). This surrogate model is learned thanks to a sequential Design of Experiments, whose points are chosen in the design space $\mathbb{X} \subset \mathbb{R}^d$, accordingly to the optimization of an acquisition criterion (see [5] for more details). The Bichon criterion [1] is a classic criteria for excursion set estimation which provides a good compromise between exploring the design space and exploiting the knowledge around the excursion set boundary.

In this work, we focus on black-box costly models with vectorial output defined by $\boldsymbol{G} := (G_1, \ldots, G_p)$. Partial excursion sets are then defined by :

$$\forall\, i \in \{1, \ldots, p\},\ \Gamma_i^\star := \{\mathbf{x} \in \mathbb{X}, G_i(\mathbf{x}) \leq T_i\}. \tag{1}$$

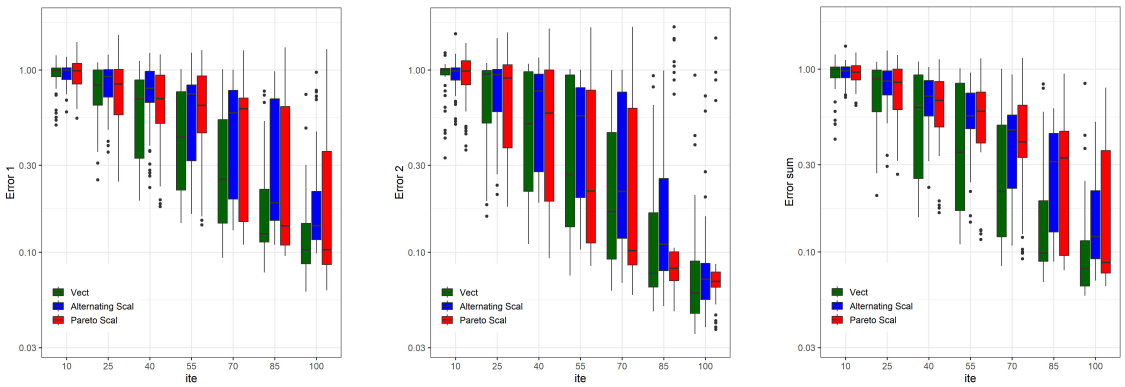
In [3], the authors proposed a criterion suitable for estimating the intersection set of partial excursion sets. In the application which motivates our study, namely a pre-calibration stage of a simulator for the design of floating wind turbines, knowledge of input values feasible for all output components is not sufficient. This is why this work aims to estimate simultaneously each partial excursion sets for each output. It enables to know for a given design space point, which output component exceeds its respective threshold.

In the first part, we introduce two natural multi-dimensional ($p = 2$) approaches based on the one-dimensional Bichon criterion. The first approach (called *Alternating Scal*) optimizes each Bichon criterion alternatively between the two components. The second approach (called *Pareto Scal*) is associated with some Pareto solutions of the bi-objective optimization problem of Bichon criteria. These two procedures use as many classic GP models as outputs (one for each component). Then, we propose a vectorial extension (called *Vect*) of the Bichon criterion based

*I benefit from an adaptation of the thesis (part-time 80% + extension of the contract) for health reasons.

on the minimum of the distances between each component of the GP and its corresponding threshold. This extension is naturally based on a multi-output GP model which takes into account the correlation between the outputs (see [4] for details) and requires calculation of orthants probabilities of multivariate normal distributions.

The different methodologies proposed above are compared on several analytical examples (with 2 then 4 input components and 2 output components), including the example below of the augmented Hartmann 4d function with 2 components and their associated thresholds $T = (-1, -1.6)$. Work on implementing these methods on the target application (application to the pre-calibration of a simulator for the design of floating wind turbines) is in progress.




(a) $\frac{\mathrm{Vol}(\hat{\Gamma}_1 \Delta \Gamma_1^\star)}{\mathrm{Vol}(\Gamma_1^\star)}$ (b) $\frac{\mathrm{Vol}(\hat{\Gamma}_2 \Delta \Gamma_2^\star)}{\mathrm{Vol}(\Gamma_2^\star)}$ (c) $\sum_i \frac{\mathrm{Vol}(\hat{\Gamma}_i \Delta \Gamma_i^\star)}{\mathrm{Vol}(\Gamma_i^\star)}$

Figure 1: Boxplots of partial relative errors (and sum), for the different criteria, in the case of 40 init. DoE of size 20 with 100 iterations, for the augmented Hartmann 4d function with $T = (-1.6, -1)$.

## Short biography (PhD student)

I'm a fourth year PhD student in AIRSEA team of Laboratoire Jean Kuntzmann at Univ. Grenoble Alpes. My thesis project initiated in October 2020 is about sampling criteria for solving scalar or vectorial excursion set estimation problems using Gaussian processes, with application to wind turbines simulator pre-calibration. This project is funded by INRIA, and is part of a collaboration with IFPEN.

## References

[1] Barron J Bichon, Michael S Eldred, Laura Painton Swiler, Sandaran Mahadevan, and John M McFarland. Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA journal*, 46(10):2459–2468, 2008.

[2] Mohamed Reda El Amri, Céline Helbert, Olivier Lepreux, Miguel Munoz Zuniga, Clémentine Prieur, and Delphine Sinoquet. Data-driven stochastic inversion via functional quantization. *Statistics and Computing*, 30:525–541, 2020.

[3] Trygve Olav Fossum, Cédric Travelletti, Jo Eidsvik, David Ginsbourger, and Kanna Rajan. Learning excursion sets of vector-valued gaussian random fields for autonomous ocean sampling. *The annals of applied statistics*, 15(2):597–618, 2021.

[4] Haitao Liu, Jianfei Cai, and Yew-Soon Ong. Remarks on multi-output gaussian process regression. *Knowledge-Based Systems*, 144:102–121, 2018.

[5] Victor Picheny, David Ginsbourger, Olivier Roustant, Raphael T Haftka, and Nam-Ho Kim. Adaptive designs of experiments for accurate approximation of a target region. 2010.