Sampling criteria for Excursion Set Estimation with Multi-Output Models

Duhamel Clément[1,3]

Mascot Num 2024

Prieur Clémentine[1,3], Helbert Céline[2],
Munoz Zuniga Miguel[4], Sinoquet Delphine[4]

April 3$^{\text{rd}}$, 2023

[1] Inria [2] ECL [3] UGA [4] IFPEN

Motivation: A pre-calibration step of a wind turbine simulator

Excursion sets to be estimated:

$$\begin{cases} \{\theta,\ \|\mathrm{Mod}_1(\theta) - \mathrm{Mod}_1^\star\| \leq T_1\} \\ \vdots \\ \{\theta,\ \|\mathrm{Mod}_p(\theta) - \mathrm{Mod}_p^\star\| \leq T_p\} \end{cases}$$

with

- θ: system parameters (stiffness coefficients),
- $\mathrm{Mod}_i(\theta)$: excited deformation mode calculated with the simulator for fixed θ,
- $\mathrm{Mod}_i^\star$: reference mode based on experimental data,
- $T_1, \ldots, T_p$: fixed thresholds.

Source: alumni.ifp-school.com

Outline

Introduction to excursion set estimation

Framework

Excursion set to estimate

$$\Gamma^\star := \Big\{ \mathbf{x} \in \mathbb{X},\ g(\mathbf{x}) \leq T \Big\},$$

with

- $\mathbb{X} \subset \mathbb{R}^d$ design space (compact),
- $g : \mathbb{X} \to \mathbb{R}$ "black-box" function (e.g. numerical simulator run),
- $T \in \mathbb{R}$ fixed threshold.

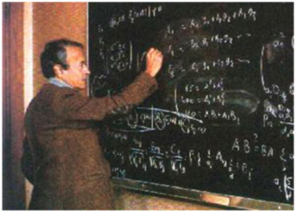
Main constraint for the estimation of $\Gamma^\star$

- To limit the number of g's expensive simulations

Gaussian Process (GP) Regression

GP Regression

- Hypothesis: black-box model g is a realization of a GP.



Professor Georges Matheron.

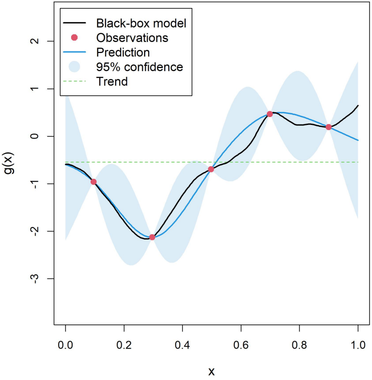
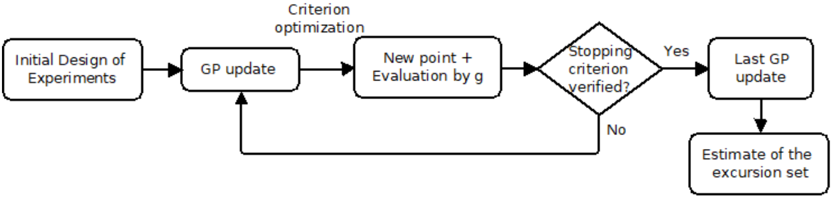


Figure 1: Example of GP Regression on 1d black-box model.

Sequential design of experiments

Sequential construction of a Design of Experiments (DoE) by GPR



Criterion choice

- Overall knowledge criteria

  ↪ Example: take the $\mathbf{x}$ that maximizes the GP model error (mse)
- Goal oriented criteria (*Picheny et al.* [2010])
  - for optimization
  - for excursion set estimation

The Bichon criterion

Notation

- $\xi(\mathbf{x})_{\mathbf{x}\in\mathbb{X}} \sim \mathrm{GP}(m,k)$: surrogate model of g,
- $\mathscr{X}_n := (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)})$: sequential DoE, $g(\mathscr{X}_n) := \left(g\left(\mathbf{x}^{(1)}\right), \dots, g\left(\mathbf{x}^{(n)}\right)\right)$,
- $\mathscr{E}_n := \left\{\xi(\mathscr{X}_n) = g(\mathscr{X}_n)\right\}$: interpolation constraints,
- $m_n(\mathbf{x}) := \mathbb{E}\left[\xi(\mathbf{x}) \middle| \mathscr{E}_n\right]$, $k_n(\mathbf{x},\mathbf{x}') := \mathrm{Cov}\left[\xi(\mathbf{x}), \xi(\mathbf{x}') \middle| \mathscr{E}_n\right]$ and $\sigma_n(\mathbf{x}) := \sqrt{k_n(\mathbf{x},\mathbf{x})}$.

Bichon criterion (*Bichon et al.* [2008])

- Oriented criterion for excursion set estimation (exploration-exploitation)

$$\mathbf{x}^{(n+1)} \in \underset{\mathbf{x}\in\mathbb{X}}{\operatorname{argmax}} \Bigg\{ \underbrace{\sigma_n(\mathbf{x})\, \mathbb{E}\left[\left(\kappa - \frac{|T - \xi(\mathbf{x})|}{\sigma_n(\mathbf{x})}\right)^+ \middle| \mathscr{E}_n\right]}_{\mathrm{EFF}(\mathbf{x})} \Bigg\},$$

with $\kappa > 0$.

Statistical interpretation

- Let $\mathbf{x}$ be fixed, suppose that $\xi(\mathbf{x}) | \mathscr{E}_n \sim \mathcal{N}(m_n(\mathbf{x}), \sigma_n(\mathbf{x}))$ with $m_n(\mathbf{x})$ unknown and $\sigma_n(\mathbf{x}) > 0$ known.
- Statistical test:

$$\mathrm{H}_0 : m_n(\mathbf{x}) = T \quad \text{against} \quad \mathrm{H}_1 : m_n(\mathbf{x}) \neq T.$$

- Conditional test statistic: $\upsilon_{\mathbf{x}} := \left(\frac{\xi(\mathbf{x}) - T}{\sigma_n(\mathbf{x})} \right) \Big| \mathscr{E}_n \sim \mathcal{N}(0,1)$ under H_0.
- We refute Hypothesis H_0 at order α if

$$|\upsilon_{\mathbf{x}}| > \kappa_\alpha,$$

  with κ_α the quantile of order $1 - \frac{\alpha}{2}$ of $\mathcal{N}(0,1)$.
- We want to select the $\mathbf{x}$ for which, on average conditioned on the event "H_0 is plausible" (i.e. $\kappa_\alpha - |\upsilon_{\mathbf{x}}| > 0$), the quantity $\kappa_\alpha - |\upsilon_x|$ is the largest possible, weighted by the probability that H_0 is plausible.
- Multiplying $\mathbb{E}\left[(\kappa_\alpha - |\upsilon_{\mathbf{x}}|)^+ \middle| \mathscr{E}_n \right]$ by $\sigma_n(\mathbf{x})$ leads to Bichon criterion.

Excursion set estimation on multi-output models

Framework

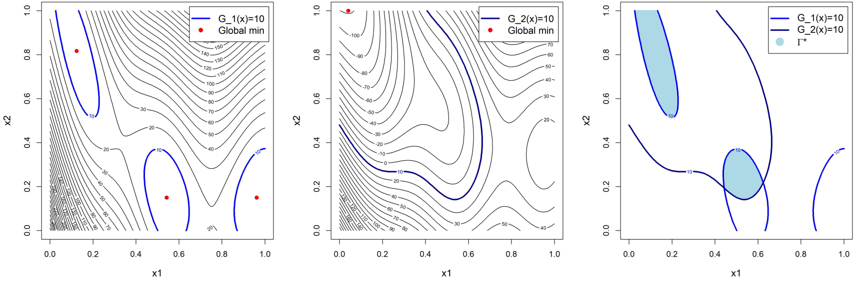
Framework

- $\mathbb{X} \subset \mathbb{R}^d$ design space (compact),
- $\boldsymbol{G} := (G_1, \ldots, G_p)^\top : \mathbb{X} \to \mathbb{R}^p$ vectorial "black-box" model,
- $\boldsymbol{T} := (T_1, \ldots, T_p)^\top \in \mathbb{R}^p$ vector of fixed thresholds,
- What are we really looking for?
  - $\forall\, i, \Gamma_i^\star := \{\mathbf{x} \in \mathbb{X}, G_i(\mathbf{x}) \leq T_i\}$
  - $\Gamma^\star := \{\mathbf{x} \in \mathbb{X}, \forall\, i, G_i(\mathbf{x}) \leq T_i\}$ $(= \cap_i \Gamma_i^\star)$

Main constraint

- Simultaneous evaluation at same $\mathbf{x}$ point in $\mathbb{X}$ (isotopic data)

A 2d example with two outputs



(a) G_1 and $\Gamma_1^\star$ (b) G_2 and $\Gamma_2^\star$ (c) $\Gamma^\star$

Figure 2: Representation of $\Gamma_i^\star$ and $\Gamma^\star$ for the Branin 2d test case with $\boldsymbol{T} := (10, 10)$.

Natural extensions of Bichon criterion (for $p = 2$)

Alternating Bichon criterion

1. $\forall i, \ \mathbf{x}_i^{(n+1)} \leftarrow$ Argmax Bichon criterion with a surrogate model on G_i and $T = T_i$
2. $\mathbf{x}^{(n+1)} \leftarrow \begin{cases} \mathbf{x}_1^{(n+1)} & \text{if } ite \text{ is even} \\ \mathbf{x}_2^{(n+1)} & \text{otherwise} \end{cases}$ with ite the iteration number

Pareto Bichon criterion

1. Obtaining the Pareto front,
2. Ideal point $I := \left(\max_j \mathrm{EFF}_1(\mathbf{x}_j), \max_j \mathrm{EFF}_2(\mathbf{x}_j)\right)$,
3. Choose the point on the Pareto front that minimizes the distance to I in $\|.\|_2$ (*Marler and Arora* [2004])

$$\mathbf{x}^{(n+1)} \in \underset{\mathbf{x}_j \in \text{Pareto}}{\operatorname{argmin}} \left\{ \sqrt{(\mathrm{EFF}_1(\mathbf{x}_j) - I_1)^2 + (\mathrm{EFF}_2(\mathbf{x}_j) - I_2)^2} \right\}.$$

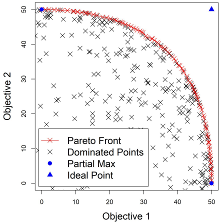


Figure 3: Pareto front example

Multidimensional GP models

Using a multidimensional GP model

- Hypothesis: model $\boldsymbol{G}$ is a realisation of a vectorial Gaussian Process $\boldsymbol{\xi} := (\xi_1, \dots, \xi_p)^\top$,
- Construction: same sequential approach as GPR,
- Challenge: Modelling the correlation between the components of $\boldsymbol{G}$

Notation

- Same as GPR by replacing g by $\boldsymbol{G}$ and ξ by $\boldsymbol{\xi}$,
- $K_n(\mathbf{x}, \mathbf{x}') := \left(\mathrm{Cov}\left[\xi_i(\mathbf{x}), \xi_j(\mathbf{x}') \middle| \mathscr{E}_n\right]\right)_{(i,j)\in\{1,\dots p\}^2}$, $\Sigma_n(\mathbf{x}) := K_n(\mathbf{x}, \mathbf{x})$ and $M_n(\mathbf{x}) = \left(M_{n,1}(\mathbf{x}), \dots, M_{n,p}(\mathbf{x})\right)^\top := \mathbb{E}\left[\boldsymbol{\xi}(\mathbf{x}) \middle| \mathscr{E}_n\right]$.

Vectorial Bichon criterion

Idea

- Generalizing Bichon criterion to a multidimensional GP model
  - taking into account the correlation between outputs.

The proposed criterion

$$\mathbf{x}^{(n+1)} := \underset{\mathbf{x}\in\mathbb{X}}{\operatorname{argmax}} \left\{ |\Sigma_n(\mathbf{x})|^{\frac{1}{p}}\, \mathbb{E}\left[\left(\kappa - \min_i \left(\frac{|T_i - \xi_i(\mathbf{x})|}{\sigma_{n,i}(\mathbf{x})} \right) \right)^+ \middle| \mathscr{E}_n \right] \right\}$$

with $\sigma_{n,i}(\mathbf{x}) := \sqrt{(\Sigma_n(\mathbf{x}))_{i,i}}$.

Explicit formulation

$$\mathbf{x}^{(n+1)} = \underset{\mathbf{x}\in\mathbb{X}}{\operatorname{argmax}} \left\{ |\Sigma_n(\mathbf{x})|^{\frac{1}{p}} \int_0^{\kappa} F_{Y_\mathbf{x}}(t)\,\mathrm{d}t \right\}$$

with $Y_\mathbf{x} := \min_i \left(\frac{|T_i - \xi_i(\mathbf{x})|}{\sigma_{n,i}(\mathbf{x})} \right)$ and $F_{Y_\mathbf{x}}$ c.d.f of $Y_\mathbf{x}$. In addition,

$$F_{Y_\mathbf{x}}(t) \overset{(p=2)}{=} \Phi(t+\alpha_1) - \Phi(-t+\alpha_1) + \Phi(t+\alpha_2) - \Phi(-t+\alpha_2) - \mathbb{P}\big((U_1, U_2) \in [\alpha_1 \pm t] \times [\alpha_2 \pm t]\big)$$

with $\alpha_i := \frac{T_i - M_{n,i}(\mathbf{x})}{\sigma_{n,i}(\mathbf{x})}$, Φ c.d.f. of $\mathcal{N}(0,1)$ and $(U_1, U_2) \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$.

Numerical experiments

A 4d example with two outputs

Augmented Hartmann 4d function (Pearson correlation $= 0.81$)

- First output component: Hartmann 4d function ($[0,1]^4 \to \mathbb{R}$) defined $\forall \mathbf{x} \in [0,1]^4$ by

$$f(\mathbf{x}) = \frac{1}{0.839} \left[1.1 - \sum_{i=1}^{4} \alpha_i \exp\left(- \sum_{j=1}^{4} A_{ij} \left(x_j - P_{ij}\right)^2 \right) \right], \qquad (1)$$

with $\alpha = (1.0, 1.2, 3.0, 3.2)^\top$,

$$\mathbf{A} = \begin{pmatrix} 10 & 3 & 17 & 3.50 \\ 0.05 & 10 & 17 & 0.1 \\ 3 & 3.5 & 1.7 & 10 \\ 17 & 8 & 0.05 & 10 \end{pmatrix} \text{ and } \mathbf{P} = 10^{-4} \begin{pmatrix} 1312 & 1696 & 5569 & 124 \\ 2329 & 4135 & 8307 & 3736 \\ 2348 & 1451 & 3522 & 2883 \\ 4047 & 8828 & 8732 & 5743 \end{pmatrix}.$$

- Second output component: modification of the Hartmann 4d function $\hookrightarrow$ (1) with $\alpha' = \alpha + \alpha_{\text{noise}}$, $A' = A + A_{\text{noise}}$, $\alpha_{\text{noise}} = \text{runif}(-1, 1, 4)$ and $A_{\text{noise}} = \text{runif}(-5, 5, 16)$ (no modification of P).
- Choice of threshold vector: $\mathbf{T} := (-1.6, -1)$,
  - Relative volume of partial excursion sets $\simeq 9.5\%$ and 5.7% for $\cap$.

Selected strategies and associated GP models

- Selected strategies
  - "Alternating Scal": SOGP + Alternating Bichon criterion
  - "Pareto Scal": SOGP + Pareto Bichon criterion
  - "Vect": MOGP + Vectorial Bichon criterion
- SOGP: Independant GPR on each output component (constant mean and Matèrn $5/2$ product kernel).
- MOGP: Multidimensional GP model (constant mean)
  - Separable kernels (*Goovaerts et al.* [1997]):

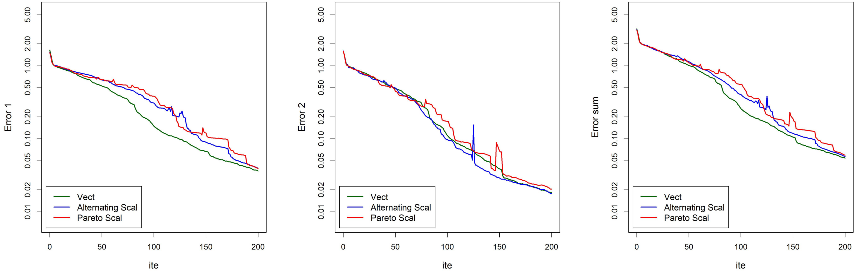
$$K\left(\mathbf{x},\mathbf{x}'\right) = \mathbf{B}\,k\left(\mathbf{x},\mathbf{x}'\right), \tag{2}$$

  - k: Matèrn $5/2$ product kernel on $\mathbb{X}$,
  - $\mathbf{B}$ spherical parametrization from *Pelamatti et al.* [2022]

$$\mathbf{B} = \sigma^2_{kOut}\begin{pmatrix} 1 & \cos(\theta_{kOut}) \\ \cos(\theta_{kOut}) & 1 \end{pmatrix} \text{ with } \sigma_{kOut} \in \mathbb{R}^+, \theta_{kOut} \in [0,\pi].$$

Means, functional boxplots and % error $< C$

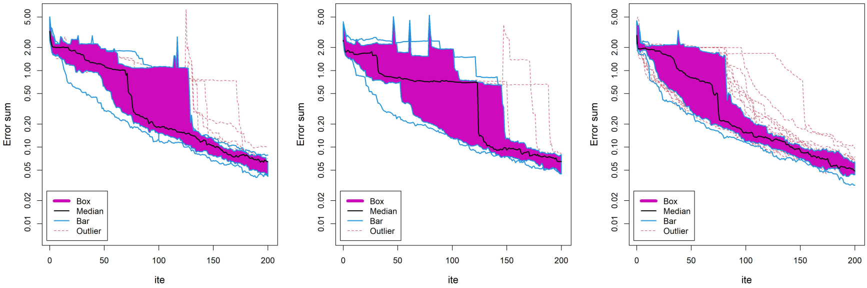
Means of partial relative approximation errors



(a) $\frac{\mathbb{P}_{\mathbb{X}}(\hat{\Gamma}_1 \Delta \Gamma_1^\star)}{\mathbb{P}_{\mathbb{X}}(\Gamma_1^\star)}$ (b) $\frac{\mathbb{P}_{\mathbb{X}}(\hat{\Gamma}_2 \Delta \Gamma_2^\star)}{\mathbb{P}_{\mathbb{X}}(\Gamma_2^\star)}$ (c) $\sum_i \frac{\mathbb{P}_{\mathbb{X}}(\hat{\Gamma}_i \Delta \Gamma_i^\star)}{\mathbb{P}_{\mathbb{X}}(\Gamma_i^\star)}$

Figure 4: Means of the partial relative errors and the sum, for the different selected strategies, in the case of enrichment of 40 LHS Maximin initial DoE of size 20 with 200 iterations, for the augmented Hartmann 4d function with $\boldsymbol{T} = (-1.6, -1)$.

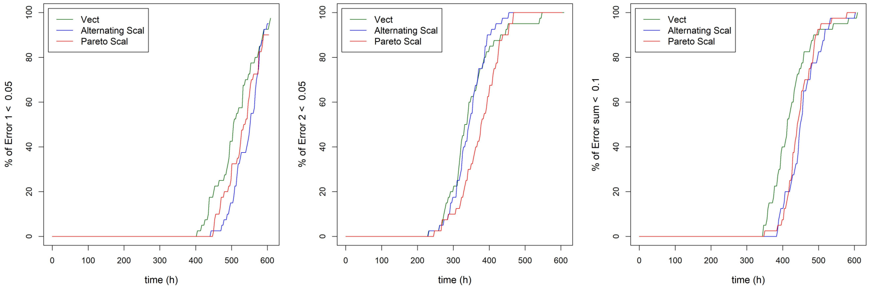
Functional boxplots of the sum of partial relative approximation errors



(a) Alternating Scal (b) Pareto Scal (c) Vect

Figure 5: Functional boxplots (MBD type) of the sum of partial relative errors, for the different selected strategies, in the case of enrichment of 40 LHS Maximin initial DoE of size 20 with 200 iterations, for the augmented Hartmann 4d function with $\boldsymbol{T} = (-1.6, -1)$.

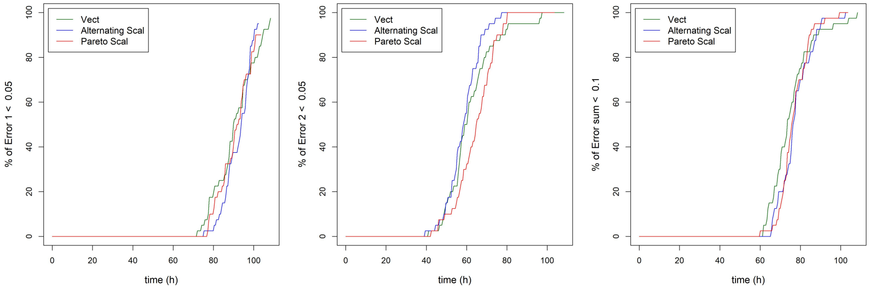
% of the relative approximation errors "definitively" $< C = 0.05\%$ (or $2C$ for the sum) (evaluation time of $g = 2$h)



(a) $\frac{\mathbb{P}_{\mathbb{X}}(\hat{\Gamma}_1 \Delta \Gamma_1^\star)}{\mathbb{P}_{\mathbb{X}}(\Gamma_1^\star)}$ (b) $\frac{\mathbb{P}_{\mathbb{X}}(\hat{\Gamma}_2 \Delta \Gamma_2^\star)}{\mathbb{P}_{\mathbb{X}}(\Gamma_2^\star)}$ (c) $\sum_i \frac{\mathbb{P}_{\mathbb{X}}(\hat{\Gamma}_i \Delta \Gamma_i^\star)}{\mathbb{P}_{\mathbb{X}}(\Gamma_i^\star)}$

Figure 6: % of (the sum of) partial relative errors "definitively" $< C$ with $C = 0.05\%$ (and 0.1% for the sum) with a 2h evaluation of g, for the different selected strategies in the case of enrichment of 40 LHS Maximin initial DoE of size 20 with 200 iterations, for the augmented Hartmann 4d function with $\boldsymbol{T} = (-1.6, -1)$

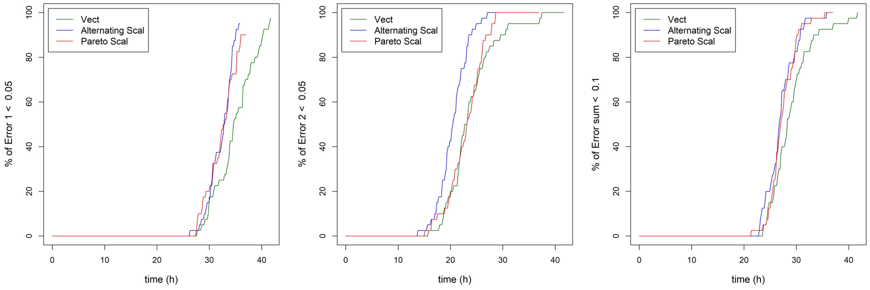
% of the relative approximation errors "definitively" $< C = 0.05\%$ (or $2C$ for the sum) (evaluation time of $g = 30$min)



(a) $\frac{\mathbb{P}_{\mathbb{X}}(\hat{\Gamma}_1 \Delta \Gamma_1^\star)}{\mathbb{P}_{\mathbb{X}}(\Gamma_1^\star)}$ (b) $\frac{\mathbb{P}_{\mathbb{X}}(\hat{\Gamma}_2 \Delta \Gamma_2^\star)}{\mathbb{P}_{\mathbb{X}}(\Gamma_2^\star)}$ (c) $\sum_i \frac{\mathbb{P}_{\mathbb{X}}(\hat{\Gamma}_i \Delta \Gamma_i^\star)}{\mathbb{P}_{\mathbb{X}}(\Gamma_i^\star)}$

Figure 7: % of (the sum of) partial relative errors "definitively" $< C$ with $C = 0.05\%$ (and 0.1% for the sum) with a 30min evaluation of g, for the different selected strategies in the case of enrichment of 40 LHS Maximin initial DoE of size 20 with 200 iterations, for the augmented Hartmann 4d function with $\boldsymbol{T} = (-1.6, -1)$

% of the relative approximation errors "definitively" $< C = 0.05\%$ (or $2C$ for the sum) (evaluation time of $g = 10$min)



(a) $\frac{\mathbb{P}_{\mathbb{X}}(\hat{\Gamma}_1 \Delta \Gamma_1^\star)}{\mathbb{P}_{\mathbb{X}}(\Gamma_1^\star)}$ (b) $\frac{\mathbb{P}_{\mathbb{X}}(\hat{\Gamma}_2 \Delta \Gamma_2^\star)}{\mathbb{P}_{\mathbb{X}}(\Gamma_2^\star)}$ (c) $\sum_i \frac{\mathbb{P}_{\mathbb{X}}(\hat{\Gamma}_i \Delta \Gamma_i^\star)}{\mathbb{P}_{\mathbb{X}}(\Gamma_i^\star)}$

Figure 8: % of (the sum of) partial relative errors "definitively" $< C$ with $C = 0.05\%$ (and 0.1% for the sum) with a 10min evaluation of g, for the different selected strategies in the case of enrichment of 40 LHS Maximin initial DoE of size 20 with 200 iterations, for the augmented Hartmann 4d function with $\boldsymbol{T} = (-1.6, -1)$

Conclusion and outlook

Conclusion

- Presentation of multidimensional excursion sets
- What are we really looking for ?
  $\hookrightarrow \forall i, \Gamma_i^\star := \{\mathbf{x} \in \mathbb{X}, G_i(\mathbf{x}) \leq T_i\}$
- Natural Criteria inspired from single-output excursion set estimation
- Multidimensional GP modelling and extension of Bichon criterion
  - Good trade-off for "Vect" in learning the 2 components

Outlook

- Extend to a more complex MOGP (not necessarily separable) with different correlation length parameters for each output
- Develop a scalar Pareto method applied to a SUR criterion, e.g. SUR Bichon (*Duhamel et al.* [2023])

Thank you for your attention !

References

Bichon B. J., Eldred M. S., Swiler L. P., Mahadevan S., & McFarland J. M. [2008]. Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA journal*, 46(10), 2459 – 2468.

Duhamel C., Helbert C., Munoz Zuniga M, Prieur, C and Sinoquet D. [2023], A SUR version of the Bichon criterion for excursion set estimation. *Statistics and Computing*, 33(2), 41.

Goovaerts P. [1997]. *Geostatistics for natural resources evaluation*. Oxford University Press on Demand.

Marler R. T. & Arora J. S. [2004]. Survey of multi-objective optimization methods for engineering. *Structural and multidisciplinary optimization*, 26, 369 – 395.

Pelamatti J., Le Riche R., Helbert C., & Blanchet-Scalliet C. [2023]. Coupling and selecting constraints in Bayesian optimization under uncertainties. *Optimization and Engineering*, 1 – 40.

Picheny V., Ginsbourger D., Roustant O., Haftka R. & Kim N-H. [2010] Adaptive Designs of Experiments for Accurate Approximation of a Target Region. *Journal of Mechanical Design*, 132(7), 071008 – 9.

Appendice

Selected strategies

- "Alternating Scal": SOGP + Alternating Bichon criterion
  - 2 optimization of $d+2$ parameters for SOGP model $+\ 2$ optimization of each partial criterion
- "Pareto Scal": SOGP + Pareto Bichon criterion
  - 2 optimization of $d+2$ parameters for SOGP model $+\ 1$ optimization on the Pareto front
- "Vect": MOGP + Vectorial Bichon criterion
  - 1 optimization of $d+3$ parameters for MOGP model $+\ 1$ optimization of vectorial criterion