On feasible set estimation with Bayesian active learning

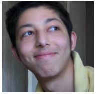
Clémentine PRIEUR

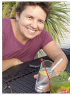
Les mobilités électriques à l'échelle d'un territoire :
approches méthodologiques et études de cas.

Séminaire scientifique - LabEx EnergyAlps
Château de Sassenage, 23 mai 2025

Clément Duhamel

Reda El Amri, Miguel Munoz Zuniga, Delphine Sinoquet





Céline Helbert

Outline

Introduction

Bayesian Active Learning

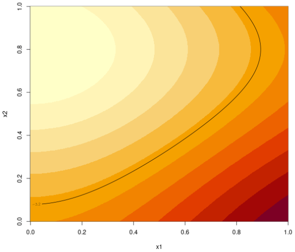
Feasible set estimation under uncertainties

Conclusion and perspectives

Let $\mathcal{X}$ and $\mathcal{Y}$ be two sets. Let $f : \mathcal{X} \to \mathcal{Y}$. Let $\mathcal{Y}_0 \subset \mathcal{Y}$. The feasible set associated to $\mathcal{Y}_0$ is defined as : $\Gamma^* := \{x \in \mathcal{X} : f(x) \in \mathcal{Y}_0\}$.

Example :

$$f : \begin{cases} [0,1]^2 \to \mathbb{R} \\ x = (x, x') \mapsto 2(1 - 2\exp(-x^2) - 1.7\exp(-2(x' - 0.8)^2)) \end{cases}$$

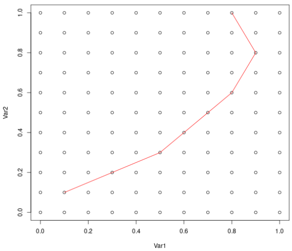


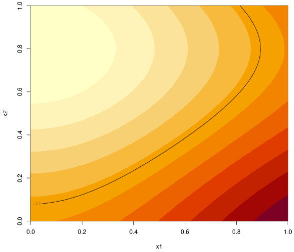
$$\Gamma^* := \{x \in \mathcal{X} : f(x) \leq -3.2\}$$

How to estimate Γ^* from model evaluations $(x_1, f(x_1)), \ldots, (x_n, f(x_n))$?

Naive approaches

1. regular grid





Drawbacks :

- curse of dimensionality (here grid of size 1001×1001 in dimension 2) ;
- arbitrariness of choosing the grid (e.g., placement).

2. random grid (e.g., uniform sampling, LHS, known priors)
   Advantage : it works in higher dimensions.
   Drawback : sampling density is not informed by the model.

Motivation 1 : pre-calibration of a wind turbine simulator

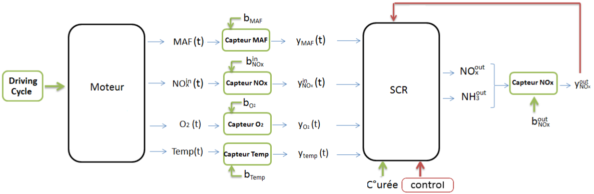
Goal

- Pre-calibrate a wind turbine simulator
  - Compare simulated modes and frequencies with experimental data (*Cadoret* [2023])

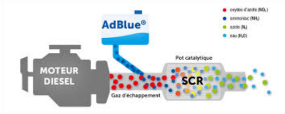
The black box model

- Inputs : 2 stiffness coefficients (Θ)
- Outputs : 13 frequencies and vibration modes $(\lambda_i(\Theta), \mathrm{Mod}_i(\Theta))$

Motivation 2 : an automotive NO_x depollution system



European emission standards : $\left\{ \begin{array}{lcl} NO_x^{out} & \leq & 80 \; mg.km^{-1} \\ NH_3^{out} & \leq & 30 \; ppm \end{array} \right.$



Outline

Introduction

Bayesian Active Learning
- Scalar output
- Vector output

Feasible set estimation under uncertainties

Conclusion and perspectives

Gaussian Process Emulator

In our setting, $f : \mathcal{X} \to \mathbb{R}^p$ is a continuous black-box model costly to evaluate and $\mathcal{X}$ is a compact subset of $\mathbb{R}^d$.

A solution is to emulate f so that we can compute approximated evaluations of f at low cost.

Gaussian prior

We assume that the deterministic black-box model f is a realization of a Gaussian random field $(Z_x)_{x \in \mathcal{X}}$ with prior mean m and covariance kernel k. Define $\Gamma := \{x \in \mathcal{X} : Z_x \in \mathcal{Y}_0\}$.

Posterior distribution

For model evaluations on a design $\mathcal{X}_n := \{x_1, \ldots, x_n\} \in \mathcal{X}^n$, denoted by $\mathbf{f}_n := \{f_1, \ldots, f_n\} \in \mathcal{Y}^n$, the posterior field, $Z \mid (Z(\mathcal{X}_n) = \mathbf{f}_n)$, is still a Gaussian process with mean and covariance kernel

$$\begin{cases} m_n(x) = m(x) + k(x, \mathcal{X}_n) k(\mathcal{X}_n, \mathcal{X}_n)^{-1} (\mathbf{f}_n - m(\mathcal{X}_n)), \\ k_n(x, y) = k(x, y) - k(x, \mathcal{X}_n) k(\mathcal{X}_n, \mathcal{X}_n)^{-1} k(\mathcal{X}_n, y),\ \sigma_n^2(x) = k(x, x). \end{cases}$$

Bayesian Active Learning

Bayesian feasible set estimation

We estimate $\Gamma^* = \{x \in \mathcal{X} : f(x) \in \mathcal{Y}_0\}$ by $\hat{\Gamma}_n = \{x \in \mathcal{X} : m_n(x) \in \mathcal{Y}_0\}$.

Sequential design of experiments

Sequentially evaluate f at points that minimize a specific acquisition criterion.

Stepwise Uncertainty Reduction

At each step, given a current design $\mathcal{X}_n$, find a new evaluation point x_{n+1} that optimally reduces the expected uncertainty on the future estimate, i.e.,

$$x_{n+1} \in \operatorname{argmin}_{x \in \mathcal{X}} \mathbb{E}\left[\mathcal{H}_{n+1}^{\texttt{uncert}}(x)\right],$$

with $\mathcal{H}_{n+1}^{\texttt{uncert}}(x)$ a so-called uncertainty measure (to be defined).

Bichon criterion [Bichon *et al.*, 2008]

Goal-oriented criterion for the DoE enrichment. It is an adaptation of EI [Jones *et al.*, 1998], introduced in the context of global optimization, to the inversion framework.

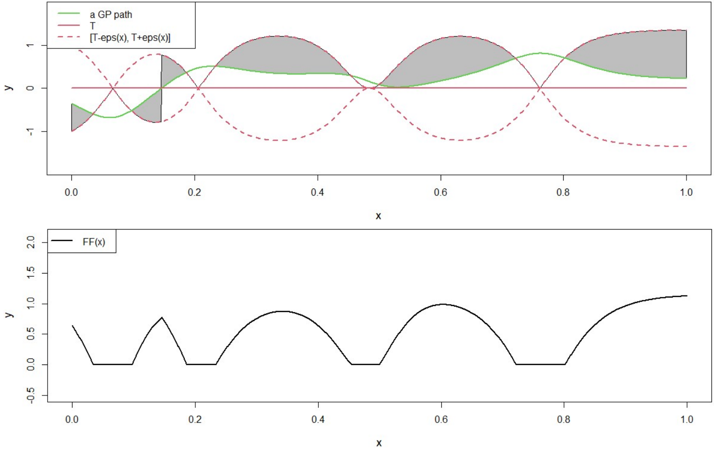
Feasibility Function

$$\begin{aligned}\mathrm{FF}(x) : &= \left(\varepsilon(x) - |c - Z_x|\right)^+ \\ &= \begin{cases} \varepsilon(x) - |c - Z_x| & \text{if } Z_x \in \left[c - \varepsilon(x), c + \varepsilon(x)\right] \\ 0 & \text{otherwise} \end{cases}\end{aligned}$$

Enrichment criterion

$x^{(n+1)} \in \underset{x \in \mathcal{X}}{\mathrm{argmax}}\, \mathrm{EFF_n}(x)$ with $\mathrm{EFF_n}(x) := \mathbb{E}\left[\mathrm{FF_n}(x) \mid (Z(\mathcal{X}_n) = \mathbf{f}_n)\right]$
where $\mathrm{FF_n}(x) := \mathrm{FF}(x)$ with $\varepsilon(x) := \kappa\sigma_n(x)$ and $\kappa > 0$.

Feasibility function



Among admissible points, add the most likely one.

SUR Bichon [Duhamel *et al.*, 2023]

Uncertainty measure

For μ a finite measure on $\mathcal{X}$ we define

$$\begin{aligned}\mathcal{H}_n^{\texttt{bichon}} &= \int_{\mathcal{X}} \mathrm{EFF_n}(z) d\mu(z) \\ &= \int_{\mathcal{X}} \mathbb{E}\Big[\mathrm{FF_n}(z) \mid (Z(\mathcal{X}_n) = \mathbf{f}_n)\Big] d\mu(z)\end{aligned}$$

and

$$\mathcal{H}_{n+1}^{\texttt{bichon}}(x) = \int_{\mathcal{X}} \mathbb{E}\Big[\mathrm{FF_{n+1}}(z) \mid (Z(\mathcal{X}_n) = \mathbf{f}_n)\,, Z_x\Big] d\mu(z).$$

Adaptive learning

$$\begin{aligned}x_{n+1} &\in \mathrm{argmin}_{x\in\mathcal{X}} \mathbb{E}\left[\mathcal{H}_{n+1}^{\texttt{bichon}}(x)\right] \\ &= \mathrm{argmin}_{x\in\mathcal{X}} \int_{\mathcal{X}} \mathbb{E}\Big[\mathrm{FF_{n+1}}(z) \mid (Z(\mathcal{X}_n) = \mathbf{f}_n)\Big] d\mu(z).\end{aligned}$$

more on computation

Numerical experiments

Model

Branin-Hoo (2D) function on $[0,1]^2$ [Picheny *et al.*, 2013].

Feasible set

$\Gamma^* := \{x \in \mathcal{X},\ f(x) \leq c\}$ with $c = 10$ so that the volume of Γ^* represents 15.74% of the total volume of $\mathcal{X}$.

Initial design

Tests are performed on 100 different initial DoEs of size 10 Maximin LHSs.

Active learning

20 iterations (1 simulation per iteration) are run. Branin function is emulated by GRF with Matérn 5/2 covariance function and constant mean.

Set estimator

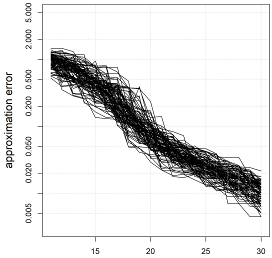
In the following, we evaluate two estimators of Γ^* :

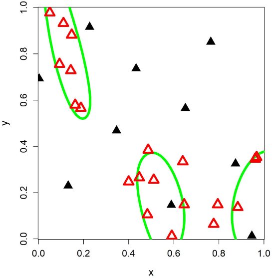
$$\hat{\Gamma}_n(\mathcal{X}_n) := \{x \in \mathcal{X} : m_n(x) \in \mathcal{Y}_0\},$$

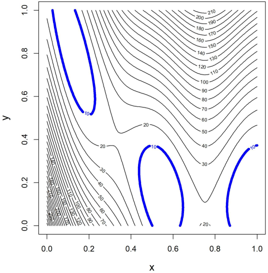
Numerical experiments

SUR Bichon $(\kappa = 1)$

$\mathrm{Err}\left(\hat{\Gamma}_n(\mathcal{X}_n)\right)$







Multioutput extension [Duhamel *et al.*, 2025]

Model

$$f = (f_1, \ldots, f_p)^T : \mathcal{X} \text{ compact subset of } \mathbb{R}^d \to \mathcal{Y} \subset \mathbb{R}^p$$

Feasible set

$$\Gamma_i^\star := \{x \in \mathcal{X}, f_i(x) \leq c_i\} \text{ , for all } 1 \leq i \leq p$$

$$\Gamma^\star := \{x \in \mathcal{X}, \forall\, 1 \leq i \leq p, f_i(x) \leq c_i\} = \cap_{i=1}^p \Gamma_i^\star)$$

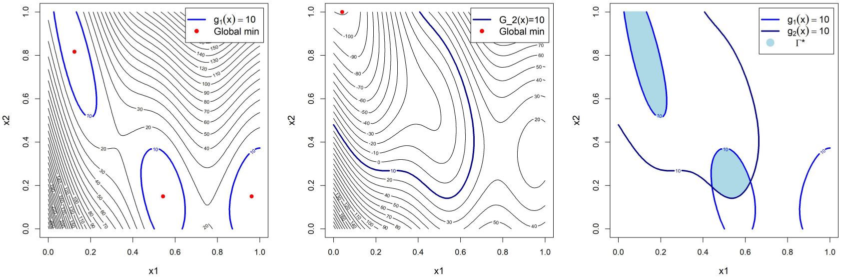
Isotopic data

at each evaluation point $x \in \mathcal{X}$, simultaneous evaluation of the p output components

Remark

For the estimation of Γ^*, vector extensions of Integrated Bernoulli Variance and Expected Measure Variance SUR criteria are proposed in [Fossum *et al.*, 2021 ; Stange and Ginsbourger, 2024].

Example with $d = 2$ and $p = 2$



$\Gamma_1^\star := \{x \in [0,1]^2, f_1(x) \leq c_1\}$ $\Gamma_2^\star := \{x \in [0,1]^2, f_2(x) \leq c_2\}$ $\Gamma^\star = \Gamma_1^* \cap \Gamma_2^*$

From now on, criteria are presented for $p = 2$ but can be generalized, at least theoretically, to any $p > 2$.

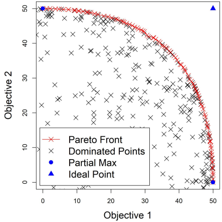
Alternating Bichon criterion

$$x_{n+1} \in \begin{cases} \underset{x \in \mathcal{X}}{\operatorname{argmax}} \, \mathrm{EFF}_{1,n}(x) & \text{if } n+1 \text{ is even} \\ \underset{x \in \mathcal{X}}{\operatorname{argmax}} \, \mathrm{EFF}_{2,n}(x) & \text{otherwise.} \end{cases}$$

Pareto Bichon criterion

$$x_{n+1} \in \underset{x \in \mathcal{X}}{\operatorname{argmin}} \left\{ \sqrt{(\mathrm{EFF}_{1,n}(x) - I_1)^2 + (\mathrm{EFF}_{2,n}(x) - I_2)^2} \right\}.$$

with $(I_1, I_2) := \left(\max_{x \in \mathcal{X}} \mathrm{EFF}_{1,n}(x), \max_{x \in \mathcal{X}} \mathrm{EFF}_{2,n}(x) \right)$ the ideal point obtained through two single-objective optimizations.



Vector output emulator

g is a realization of a vector Gaussian random field $Z := (Z_1, \dots, Z_p)^\top$,

Notation

$$K_n(x, x') := \left(\mathrm{Cov}\big(Z_i(x), Z_j(x') \mid (Z(\mathcal{X}_n) = \mathbf{f}_n)\big), (i,j) \in \{1, \dots p\}^2\right)$$

$$\Sigma_n(x) := K_n(x, x) \text{ and}$$

$$M_n(x) = \big(M_{n,1}(x), \dots, M_{n,p}(x)\big)^\top := \mathbb{E}\big(Z_x \mid (Z(\mathcal{X}_n) = \mathbf{f}_n)\big).$$

kernel choice

Output extension of Bichon criterion

$$x_{n+1} \in \operatorname*{argmax}_{x \in \mathcal{X}} \mathrm{VEFF_n}(x)$$

with $\mathrm{VEFF_n}(x) := \det\big(\Sigma_n(x)\big)^{\frac{1}{2p}} \mathbb{E}\left[\left(\kappa - \min_i \left(\frac{|c_i - Z_{i,x}|}{\sigma_{n,i}(x)}\right)\right)^+ \mid (Z(\mathcal{X}_n) = \mathbf{f}_n)\right]$

and $\sigma_{n,i}(x) := \sqrt{(\Sigma_n(x))_{i,i}}$. more on computation

Numerical experiments

Performance criteria

For $i = 1, 2$,

$$\mathrm{Err}_i := \frac{\mu(\hat{\Gamma}_i \Delta \Gamma_i^\star)}{\mu(\Gamma_i^\star)} \text{ with } \hat{\Gamma}_i := \left\{ x \in \mathcal{X},\ M_{n,i}(x) \leq c_i \right\}.$$

We also compute :

$$\mathrm{Err}_{\mathsf{sum}} := \mathrm{Err}_1 + \mathrm{Err}_2.$$

Model, $d = 2$ and $p = 2$

2D-Branin function

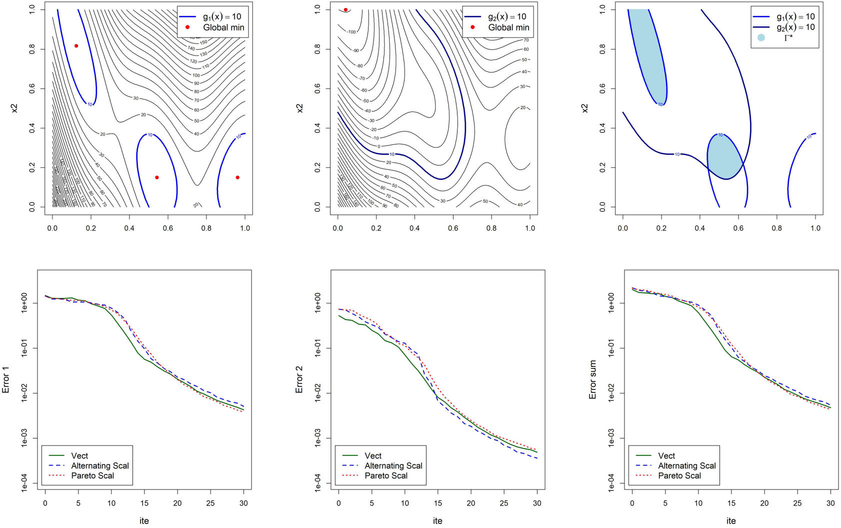
$$g_1(x) := \left(\bar{x}_2 - \frac{5\bar{x}_1^2}{4\pi^2} + \frac{5\bar{x}_1}{\pi} - 6 \right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(\bar{x}_1) + 10$$

and a slightly modified version of it

$$g_2(x) := \left(\bar{x}_2 - \frac{3\bar{x}_1^2}{4\pi^2} + \frac{4\bar{x}_1}{\pi} - 6 \right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(\bar{x}_1) + 2\bar{x}_1 - 9\bar{x}_2 + 32$$

with $\bar{x}_1 := 15x_1 - 5$ and $\bar{x}_2 := 15x_2$.

Feasible sets for $c_1 = c_2 = 10$, $\kappa = 1$. Errors are plotted as average on 40 initial Maximin LHSs of size 5 and with 30 enrichment iterations.



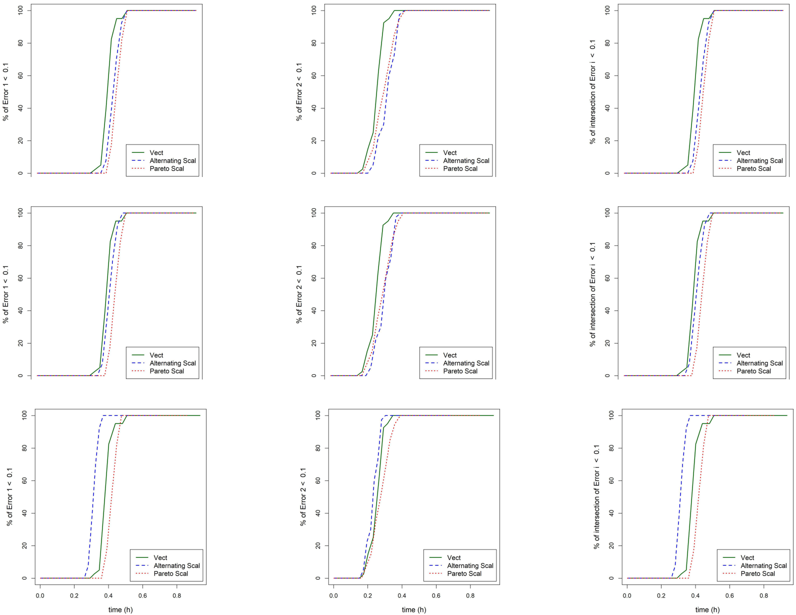
Taking computational cost into account

$$\mathrm{DP}_i(t,K) \quad := \quad 100 \times \frac{\#\left\{j \in \{1,\dots,\mathrm{N}_{\mathrm{Repet}}\}, \forall\, \tilde{t} \geq t,\, \mathrm{Err}_i^{(j)}(\tilde{t}) < K\right\}}{\mathrm{N}_{\mathrm{Repet}}},$$

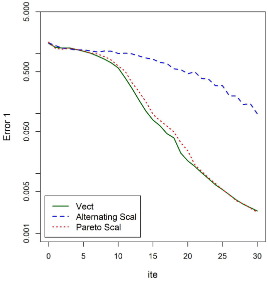
$$\mathrm{DP}_{\mathrm{tot}}(t,K) \quad := \quad 100 \times \frac{\#\left\{j \in \{1,\dots,\mathrm{N}_{\mathrm{Repet}}\}, \forall\, i, \forall\, \tilde{t} \geq t,\, \mathrm{Err}_i^{(j)}(\tilde{t}) < K\right\}}{\mathrm{N}_{\mathrm{Repet}}},$$

with $\mathrm{N}_{\mathrm{Repet}}$ the number of repetitions.

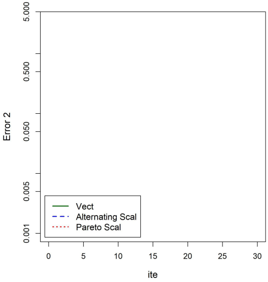
We choose $c = (10, 10)$. Data profiles are plotted with $K = 10\%$ for evaluation times equal to 3h (top line), 10min (middle line) and 1min (bottom line), respectively, and with $\mathrm{N}_{\mathrm{Repet}} = 40$ initial Maximin LHSs of size 5 on 30 iterations.


% of Error 1 < 0.1
% of Error 2 < 0.1
% of intersection of Error i < 0.1
Vect
Alternating Scal
Pareto Scal
time (h)

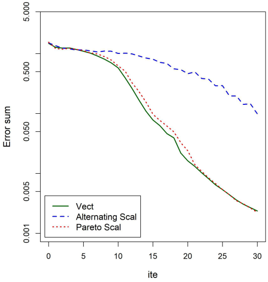
Example for which alternating Bichon criterion fails : $c_1 = 10$, $c_2 = 10^4$



(a) Err_1



(b) Err_2



(c) $\mathrm{Err}_{\text{sum}}$

Motivation 1 : pre-calibration of a wind turbine simulator

Goal

- Pre-calibrate a wind turbine simulator
  - Compare simulated modes and frequencies with experimental data (*Cadoret* [2023])

The black box model

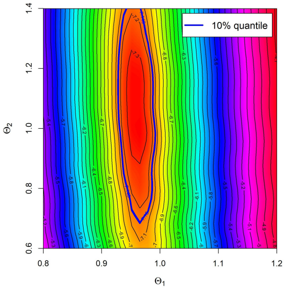
- Inputs : 2 stiffness coefficients (Θ)
- Outputs : 13 frequencies and vibration modes $(\lambda_i(\Theta), \mathrm{Mod}_i(\Theta))$

Define

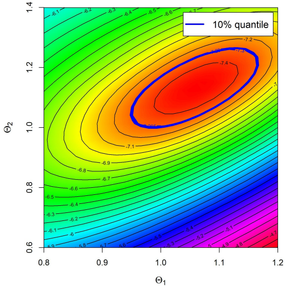
$$\Gamma_i^\star := \Big\{ \Theta \in \mathbb{X},\ g_i(\Theta) \leq T_i \Big\},\ \text{for } i \in \{1,2\}$$

with

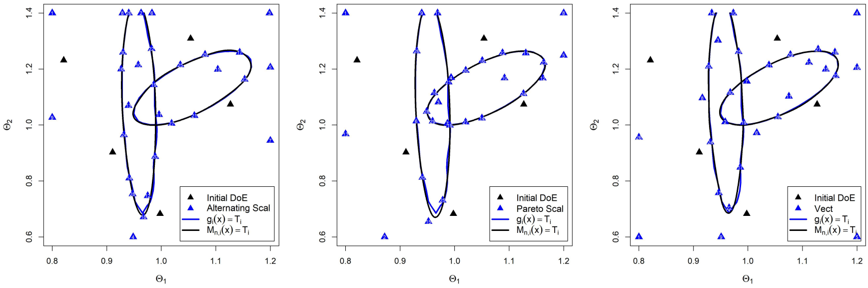
- $\mathbb{X} := [0.8, 1.2] \times [0.6, 1.4]$: design space, and T_i : thresholds,
- $g_i(\Theta) := \ln\left(\frac{\mathrm{Meas}_i(\Theta)}{1-\mathrm{Meas}_i(\Theta)}\right)$ and $\mathrm{Meas}_i(\Theta) := \left(1 - \frac{|<\mathrm{Mod}_i^\star, \mathrm{Mod}_i(\Theta)>|^2}{\|\mathrm{Mod}_i^\star\|^2 \|\mathrm{Mod}_i(\Theta)\|^2}\right)$.



(a) $g_1(\Theta)$



(b) $g_2(\Theta)$



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

Enrichment after 30 iterations from an initial LHS Maximin DoE (size 5) for different criteria, in the pre-calibration with two main modes with $\boldsymbol{T} = (7.148, 7.296)$. Partial excursion set boundaries from a 30×30 grid are overlaid.

Outline

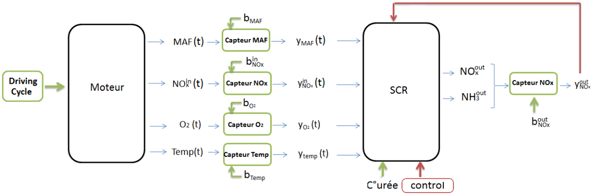
Introduction

Bayesian Active Learning

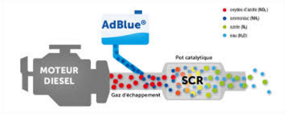
Feasible set estimation under uncertainties

Conclusion and perspectives

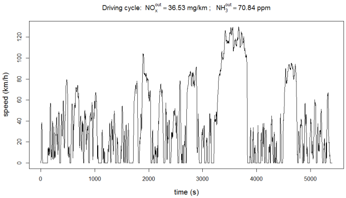
Motivation 2 : an automotive $\mathrm{NO_x}$ depollution system

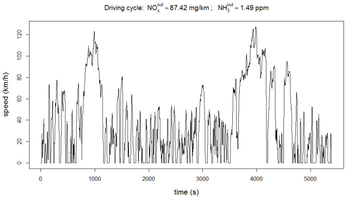


European emission standards : $\begin{cases} \mathrm{NO_x^{out}} & \leq & 80 \; mg.km^{-1} \\ \mathrm{NH_3^{out}} & \leq & 30 \; ppm \end{cases}$

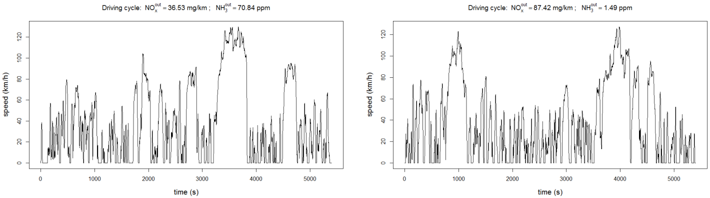


With the most recent rules, robust procedures for the emission testing of vehicles in real driving are required.

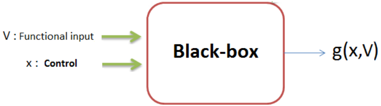




With the most recent rules, robust procedures for the emission testing of vehicles in real driving are required.



Mathematical formulation (1/2)



$$g \quad : \quad \mathcal{X} \times \mathcal{V} \to \mathbb{R}$$
$$(x, \mathbf{v}) \mapsto g(x, \mathbf{v})$$

where $\mathcal{X} \subset \mathbb{R}^d$ is compact and $\mathcal{V}$ a functional space.

Probabilistic description of uncertainty : $\mathbf{V}$ is a random variable valued in $\mathcal{V}$.

Mathematical formulation (2/2)

For a fixed $t \in \mathbb{R}$, define

$$\Gamma^*_{a.s.} := \{x \in \mathcal{X} \ s.t. \ g(\mathbf{x}, \mathbf{V}) \leq c \text{ almost surely}\},$$

$$\Gamma^*_{\alpha} := \{x \in \mathcal{X} \ s.t. \ \mathbb{P}(g(x, \mathbf{V}) \leq c) \ \geq \ 1 - \alpha\},$$

$$\Gamma^* := \{x \in \mathcal{X} \ s.t. \ f(x) = \mathbb{E}[g(x, \mathbf{V})] \leq c\} = g^{-1}(\mathcal{Y}_0), \text{where } \mathcal{Y}_0 = (-\infty, c].$$

Objective :
estimate $\Gamma^* \subset \mathbb{R}^d$ from model evaluations.

Note that the estimation of Γ^*_{α} is related to Quantile Set Inversion (see, e.g., [Ait Abdelmalek-Lomenech *et al.*]).

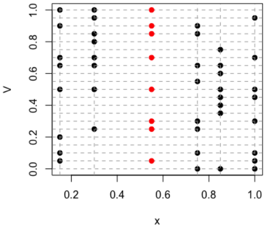
Issues :

- each evaluation $f(x)$ requires the estimation of $\mathbb{E}[g(x, \mathbf{V})]$,
- $\mathbf{V}$ is known through a set of κ realizations $\Xi = \{\mathbf{v}_1, \dots \mathbf{v}_{\kappa}\}$,
- each evaluation $g(x, \mathbf{v})$ is costful.

We proposed two different strategies to solve the problem.

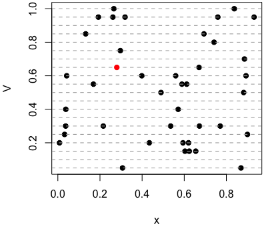
Strategy I [El Amri *et al.*, 2020]

- build a metamodel for f and choose $x_{n+1} \in \mathcal{X}$,
- estimate $f(x_{n+1}) = \mathbb{E}[g(x_{n+1}, \mathbf{V})]$ with ℓ evaluations of $g(x_{n+1}, \cdot)$ selected by quantization.



Strategy II [El Amri *et al.*, 2023]

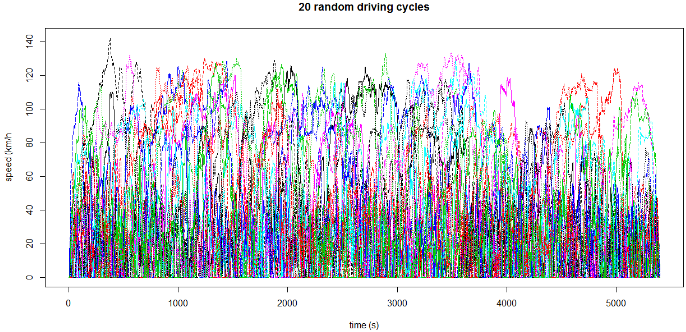
- build a metamodel for g,
- choose $(x_{n+1}, \mathbf{v}_{n+1}) \in \mathcal{X} \times \Xi$, evaluate g at this new point.



IFPEN test case : control strategy for an automotive NO_x depollution system

- $\mathrm{NH}_3^{\mathrm{out}} \leq 30$ *ppm*,
- 2 control parameters in $\mathcal{X} = [0, 0.6] \times [0, 0.6]$,
- **V** represented by 100 driving cycles on $[0, 5400s]$,

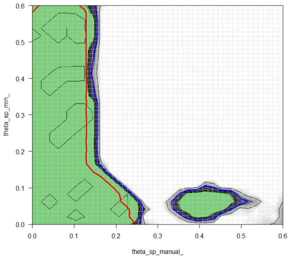
- Strategy I : initial DoE 8 points in $\mathcal{X} \subset \mathbb{R}^2$ (a mean of 23 calls to g at each point)
- Strategy II : initial DoE 5*22 in $\mathcal{X} \subset \mathbb{R}^{2+20}$ (20 eigenfunctions for 80% explained variance)



IFPEN test case : control strategy for an automotive NO_x depollution system

- $\mathrm{NH}_3^{\mathrm{out}} \leq 30$ *ppm*,
- 2 control parameters in $\mathcal{X} = [0, 0.6] \times [0, 0.6]$,
- **V** represented by 100 driving cycles on $[0, 5400s]$,

$n =$ **510** calls to g
Strategy I versus Strategy II



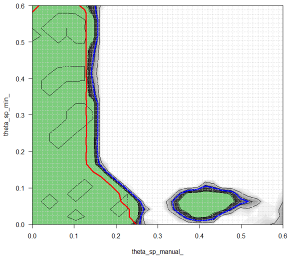
- Strategy I : initial DoE 8 points in $\mathcal{X} \subset \mathbb{R}^2$ (a mean of 23 calls to g at each point)
- Strategy II : initial DoE 5*22 in $\mathcal{X} \subset \mathbb{R}^{2+20}$ (20 eigenfunctions for 80% explained variance)

IFPEN test case : control strategy for an automotive NO_x depollution system

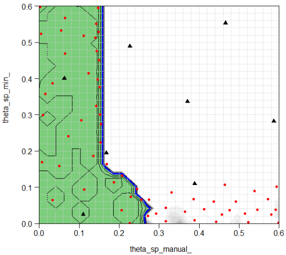
- $\mathrm{NH}_3^{\mathrm{out}} \leq 30$ *ppm*,
- 2 control parameters in $\mathcal{X} = [0, 0.6] \times [0, 0.6]$,
- **V** represented by 100 driving cycles on $[0, 5400s]$,

- Strategy I : initial DoE 8 points in $\mathcal{X} \subset \mathbb{R}^2$ (a mean of 23 calls to g at each point)
- Strategy II : initial DoE 5*22 in $\mathcal{X} \subset \mathbb{R}^{2+20}$ (20 eigenfunctions for 80% explained variance)

$n = \mathbf{510}$ calls to g
Strategy I versus Strategy II



Reference excursion set : $n = \mathbf{1575}$ calls to g with Strategy I



Outline

Introduction

Bayesian Active Learning

Feasible set estimation under uncertainties

Conclusion and perspectives

Conclusion

- Bayesian Active Learning is a powerfull tool for feasible set estimation.
- We were able to handle :
  - vector output models,
  - estimation under uncertainties.

Perspectives

- to prove convergence results for a regularized version of SUR Bichon criterion,
- to consider more general covariance structures for MOGP,
- to develop efficient implementation for more than two outputs,
- to extend Bichon criterion for estimation under uncertainties,
- to estimate feasible sets with very small relative volume (rare event), . . .

MIAI Chair (just accepted)

BALTEEC : Bayesian Active Learning Techniques for Energy EfficienCy in buildings

Thanks for your attention !

Some references I

Ait Abdelmalek-Lomenech, R., Bect, J., Chabridon, V., and Vazquez, E. (2024).
Bayesian sequential design of computer experiments for quantile set inversion.
Technometrics, pages 1–10.

Bichon, B. J., Eldred, M. 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.

Chevalier, C. (2013).
Fast uncertainty reduction strategies relying on Gaussian process models.
PhD thesis.

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 *mODa 10–Advances in Model-Oriented Design and Analysis : Proceedings of the 10th International Workshop in Model-Oriented Design and Analysis Held in Łagów Lubuski, Poland, June 10–14, 2013*, pages 35–43. Springer.

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.

Some references II

Duhamel, C., H. C. M. Z. M. P. C. and Sinoquet, D. (2025).
Active learning strategies for the estimation of a feasible set dened from a vector output black-box simulator.
https://hal.science/hal-04970769.

El Amri, M. R., Helbert, C., Lepreux, O., Zuniga, M. M., Prieur, C., and Sinoquet, D. (2020).
Data-driven stochastic inversion via functional quantization.
Statistics and Computing, 30(3) :525–541.

El Amri, M. R., Helbert, C., Zuniga, M. M., Prieur, C., and Sinoquet, D. (2023).
Feasible set estimation under functional uncertainty by gaussian process modelling.
Physica D : Nonlinear Phenomena, 455 :133893.

Fossum, T. O., Travelletti, C., Eidsvik, J., Ginsbourger, D., and Rajan, K. (2021).
Learning excursion sets of vector-valued gaussian random fields for autonomous ocean sampling.
The annals of applied statistics, 15(2) :597–618.

Goovaerts, P. (1997).
Geostatistics for natural resources evaluation.
Oxford University Press, USA.

Janusevskis, J. and Le Riche, R. (2013).
Simultaneous kriging-based estimation and optimization of mean response.
Journal of Global Optimization, 55(2) :313–336.

Some references III

Jones, D. R., Schonlau, M., and Welch, W. J. (1998).
Efficient global optimization of expensive black-box functions.
Journal of Global optimization, 13(4) :455–492.

Molchanov, I. S. (2005).
Theory of random sets, volume 87.
Springer.

Paciorek, C. J. (2003).
Nonstationary Gaussian processes for regression and spatial modelling.
PhD thesis, Carnegie Mellon University.

Pelamatti, J., Le Riche, R., Helbert, C., and Blanchet-Scalliet, C. (2024).
Coupling and selecting constraints in bayesian optimization under uncertainties.
Optimization and Engineering, 25(1) :373–412.

Picheny, V., Wagner, T., and Ginsbourger, D. (2013).
A benchmark of kriging-based infill criteria for noisy optimization.
Structural and Multidisciplinary Optimization, 48(3) :607–626.

Stange, P. and Ginsbourger, D. (2024).
Consistency of some sequential experimental design strategies for excursion set estimation based on vector-valued gaussian processes.
Electronic Journal of Statistics, 18(2) :5091–5131.

Computation

$$\mathbb{E}\Big[\big(\kappa\sigma_{n+1}(z) - |c - Z_z|\big)^+ \mid (Z(\mathcal{X}_n) = \mathbf{f}_n)\Big]$$

$$= (m_n(z) - c)\left[2\,\phi\left(\frac{c - m_n(z)}{\sigma_n(z)}\right) - \phi\left(\frac{c^- - m_n(z)}{\sigma_n(z)}\right) - \phi\left(\frac{c^+ - m_n(z)}{\sigma_n(z)}\right)\right]$$

$$- \sigma_n(z)\left[2\,\varphi\left(\frac{c - m_n(z)}{\sigma_n(z)}\right) - \varphi\left(\frac{c^- - m_n(z)}{\sigma_n(z)}\right) - \varphi\left(\frac{c^+ - m_n(z)}{\sigma_n(z)}\right)\right]$$

$$+ \kappa\sigma_{n+1}(z)\left[\phi\left(\frac{c^+ - m_n(z)}{\sigma_n(z)}\right) - \phi\left(\frac{c^- - m_n(z)}{\sigma_n(z)}\right)\right]$$

where $c^\pm := c \pm \kappa\sigma_{n+1}(z)$, φ and ϕ denote the probability density and cumulative distribution function of the standard normal distribution, respectively.

Then, the integral in the active learning criterion is computed with $\mathrm{n.points} = 10^4$. go back

Let $V_{x,n} := \min_{1\leq i\leq p}\left(\frac{|c_i - Z_{i,x}|}{\sigma_{n,i}(x)}\right) \mid (Z(\mathcal{X}_n) = \mathbf{f}_n)$.

Explicit formulation

go back

$$\mathrm{VEFF}(x) = \det\big(\Sigma_n(x)\big)^{\frac{1}{2p}} \int_0^{\kappa} F_{V_{x,n}}(t)\, dt.,$$

with $F_{V_{x,n}}$ the cumulative distribution function of $V_{x,n}$.

Explicitation for $p = 2$

$$\begin{aligned} F_{V_{x,n}}(t) &= \phi(t + \alpha_{1,n}) - \phi(-t + \alpha_{1,n}) + \phi(t + \alpha_{2,n}) - \phi(-t + \alpha_{2,n}) \\ &\quad - \mathbb{P}\big((U_1, U_2) \in [\alpha_1 - t, \alpha_1 + t] \times [\alpha_2 - t, \alpha_2 + t]\big), \end{aligned}$$

where $\alpha_{i,n} := \frac{c_i - M_{n,i}(x)}{\sigma_{n,i}(x)}$, ϕ c.d.f. of $\mathcal{N}(0,1)$ and

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_n \\ \rho_n & 1 \end{pmatrix}\right)$$

with ρ_n the correlation coefficient between $Z_{1,x}$ and $Z_{2,x}$ cond. on $Z(\mathcal{X}_n) = \mathbf{f}_n$.

Kernel choice go back

We choose an ICM-type separable kernel [Goovaerts *et al.*, 1997] :

$$\left(K(x,x')_{i,j}\right)_{1\leq i,j\leq p} := k(x,x')\,\mathbf{B},$$

with k a scalar kernel on $\mathcal{X}^2$ and $\mathbf{B}$ a $p \times p$ symmetric positive-definite matrix.

For $p = 2$ we consider, as in [Pelamatti *et al.*, 2024] ,

$$\mathbf{B} = \sigma^2_{kOut} \begin{pmatrix} 1 & \cos(\theta_{kOut}) \\ \cos(\theta_{kOut}) & 1 \end{pmatrix}$$

with $\sigma_{\mathrm{kOut}} \in \mathbb{R}$ a scaling factor common to all components (homoscedasticity), and $\theta_{kOut} \in [0, \pi]$ the parameter controlling output correlations under spherical parameterization. We choose k as :

$$k(x,x') := \sigma^2_{\text{common}} \prod_{j=1}^{d} R_{\mathrm{Matérn}\,5/2}\left(|x_j - x'_j|, \theta_j\right), \quad \forall\,(x,x') \in \mathcal{X}^2$$

with $\sigma^2_{\text{common}} = 1$. Hyperparameters are estimated by maximum likelihood.