



Bayesian multi-objective optimization for quantitative risk assessment in microbiology

Subhasish Basak, Julien Bect, Laurent Guillier, Fanny Tenenhaus-Aziza, Janushan Christy, Emmanuel Vazquez

► To cite this version:

Subhasish Basak, Julien Bect, Laurent Guillier, Fanny Tenenhaus-Aziza, Janushan Christy, et al.. Bayesian multi-objective optimization for quantitative risk assessment in microbiology. MASCOT-NUM 2022, Jun 2022, Clermont-Ferrand, France. . hal-03715857

HAL Id: hal-03715857

<https://hal.science/hal-03715857>

Submitted on 6 Jul 2022

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

BAYESIAN MULTI-OBJECTIVE OPTIMIZATION FOR QUANTITATIVE RISK ASSESSMENT IN MICROBIOLOGY

Subhasish Basak^{1,2}, Julien Bect², Laurent Guillier¹, Fanny Tenenhaus-Aziza³, Janushan Christy⁴, Emmanuel Vazquez²

¹Agence nationale de sécurité sanitaire (ANSES), 94700, Maisons-Alfort, France

²Université Paris-Saclay, CNRS, CentraleSupélec, L2S, 91190, Gif-sur-yvette, France

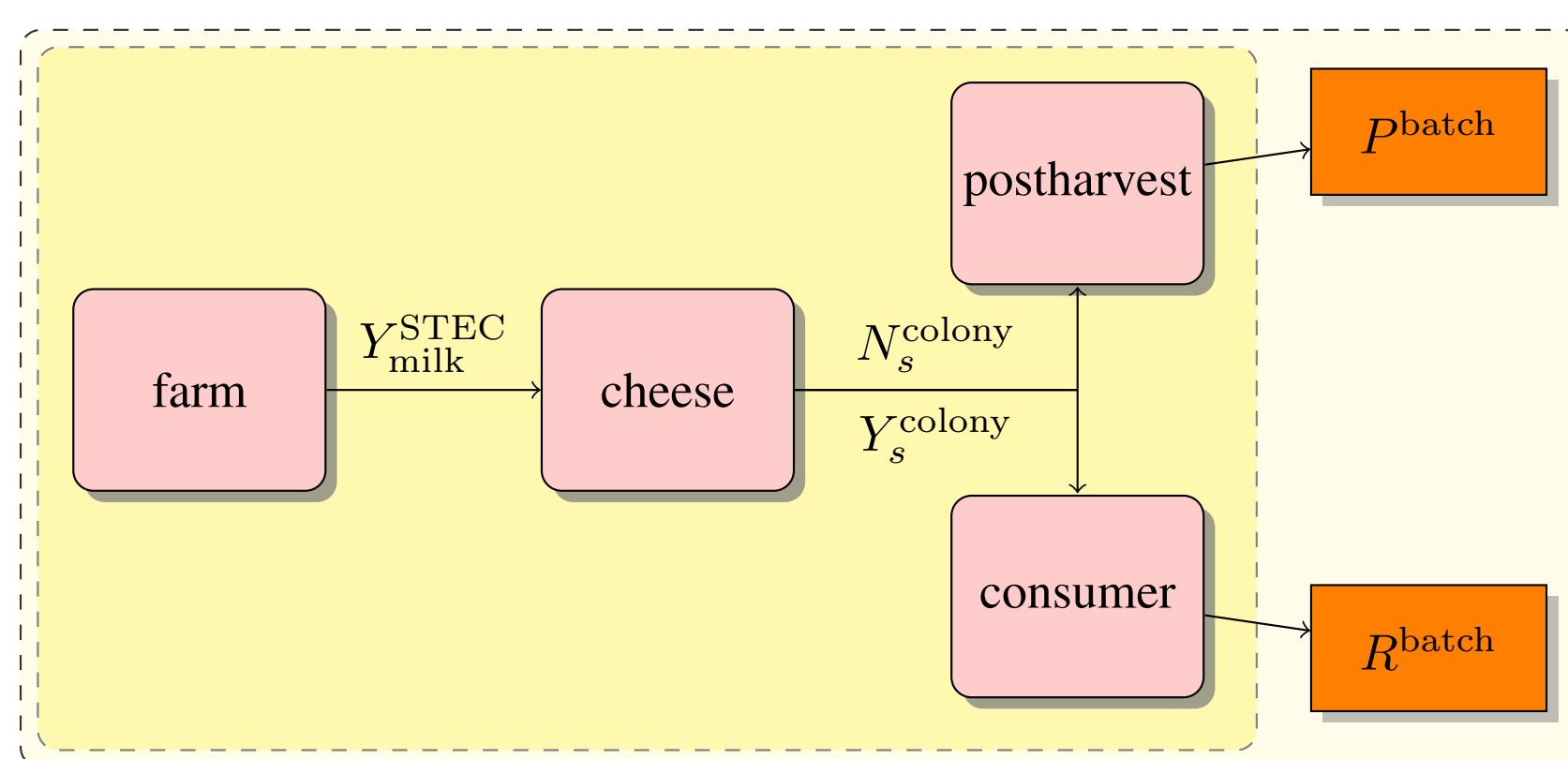
³Centre national interprofessionnel de l'économie laitière (CNIEL), 75009, Paris, France

⁴Centre technique d'expertise agroalimentaire (ACTALIA), 74800, La-Roche-sur-Foron, France

Context

As a part of the European project [ArtiSaneFood](#), the primary goal of this collaborative work between ANSES, CNIEL and L2S is to establish efficient bio-intervention strategies for cheese producers in France, in order to “economically” reduce the risk of Haemolytic Uremic Syndrome (HUS) caused by Shiga-Toxin producing Escherichia Coli (STEC) present in raw-milk soft cheese. This translates into a multi-objective optimization problem of a stochastic simulator based on a quantitative risk assessment (QRA) model proposed by Perrin et al. (2014), to estimate the Pareto optimal solutions for the process intervention parameters.

QRA simulator



Batch level stochastic simulator

The farm-to-fork continuum for one batch

• Farm module

STEC and E.coli strains follow same fecal route!

– A batch of milk is tested with probability $p_{\text{test}}^{\text{milk}}$

– Farms with E.coli conc. $> l^{\text{sort}}$ are rejected

STEC conc. $Y_{\text{milk}}^{\text{STEC}}$ in farm milk is computed

• Cheese module

STEC cells form colonies (clusters) inside cheese

– No. of colonies (Poisson): N_s^{colony}

– Size of colonies (LogNormal): Y_s^{colony} for strain s

• Consumer module

Batch risk is computed using a dose-response model:

$$\Gamma = \sum_s N_s^{\text{colony}} \cdot Y_s^{\text{colony}} \quad (1)$$

$$R^{\text{batch}} = \sum_{\text{age}} g_{\text{age}} \int_{\Gamma} P[\text{HUS}|\gamma, \text{age}] \cdot p(\gamma) d\gamma \quad (2)$$

averaging over consumer age

• Post-harvest module

Proportion of rejected batches is estimated using number of test sample n_{sample} given a test rate $p_{\text{test}}^{\text{cheese}}$

$$P^{\text{batch}} = P[\Gamma > 0] = 1 - \exp(-K \cdot n_{\text{sample}})$$

Quantities of Interest (QoI)

Several batches are simulated to estimate

$$R_{\text{avg}} = \mathbb{E}[R^{\text{batch}} \cdot (1 - P^{\text{batch}} \cdot p_{\text{test}}^{\text{cheese}})]$$

$$P_{\text{avg}} = \mathbb{E}[P^{\text{batch}} \cdot p_{\text{test}}^{\text{cheese}}]$$

QoIs are

$$\bullet \text{Relative risk: } f_1 = \frac{R_{\text{avg}}}{(1 - P_{\text{avg}}) \cdot K_1}, (K_1: \text{baseline risk})$$

$$\bullet \text{Batch rejection rate: } f_2 = P_{\text{avg}}$$

Multi-objective optimization

We consider the multi-objective optimization problem

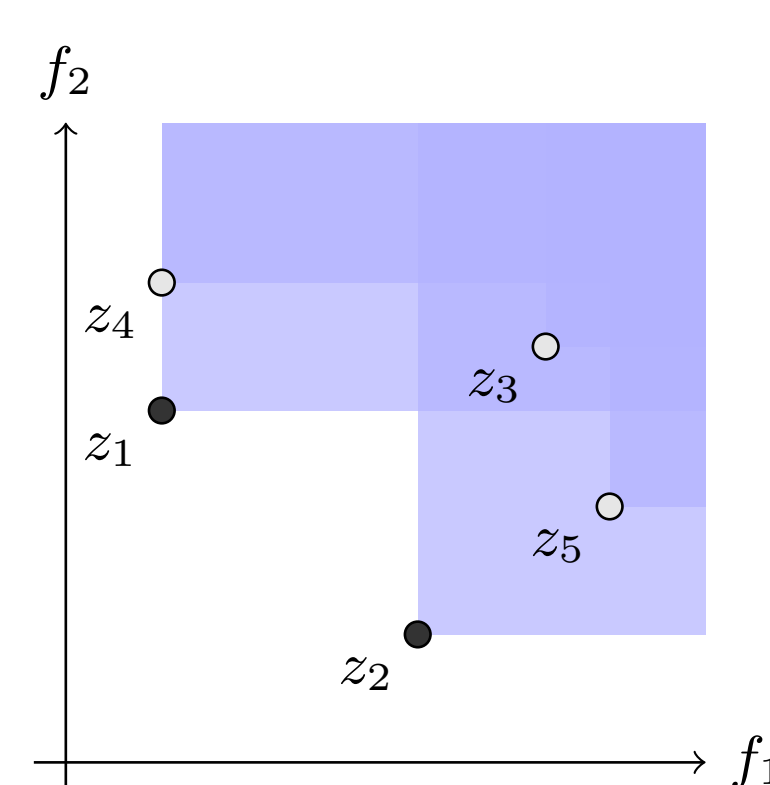
$$\min_{x \in \mathbb{X}} f(x)$$

where, $f = (f_1, f_2, \dots)$

- The solution set $\mathcal{P}$ contains **Pareto optimal** points

$$\mathcal{P} = \{x \in \mathbb{X} : \nexists x' \in \mathbb{X}, f(x') \prec f(x)\}$$

where $f' \prec f \implies f'_i \leq f_i, \forall i$, with at least one of the inequalities being strict



Pareto optimal points z_1 and z_2

- In a stochastic setting, we assume additive noise: for a given $x_i \in \mathbb{X}$, we observe $Z_i = f(x_i) + \varepsilon_i$, $\varepsilon_i \sim \mathcal{N}(0, \Sigma)$

PALS

Proposed by Zuluaga et al. (2013) and extended for the **stochastic** case by Barracosa et al. (2021).

- Initial design $I_0, n = 0, \beta > 0$

- **Step I** – $\forall x \in \mathbb{X}$, construct confidence rectangles

$$R_n(x) = \{z \in \mathbb{R}_+^2 : R_n^-(x) \prec z \prec R_n^+(x)\}$$

using posterior means $\mu_{n,j}(x)$ and variances $\sigma_{n,j}^2(x)$

$$R_n^\pm(x) = \mu_{n,j}(x) \pm \beta^{1/2} \sigma_{n,j}(x)$$

based on I_n (indep. GP models for $f_j, j \in \{1, 2\}$)

- **Step II** – Classify each $x \in \mathbb{X}$ into one of three classes

– **Deemed Pareto optimal** P_n :

$$P_n = \{x \in \mathbb{X} | \nexists x' \in \mathbb{X} \setminus \{x\}, R_n^-(x') \prec R_n^+(x)\}$$

– **Non Pareto optimal** N_n :

$$N_n = \{x \in \mathbb{X} | \exists x' \in \mathbb{X} \setminus \{x\}, R_n^+(x') \prec R_n^-(x)\}$$

– **Unclassified** $U_n = \mathbb{X} \setminus (P_n \cup N_n)$

- **Step III** – Sample the next point of evaluation

$$X_{n+1} = \arg \max_{x \in (U_n \cup P_n)} \|R_n^-(x) - R_n^+(x)\|_2$$

$$I_{n+1} = I_n \cup (X_{n+1}, Z_{n+1}), Z_{n+1} = f(X_{n+1}) + \varepsilon_{n+1}$$

- Increase n and **Repeat I-III** until $n \leq n_{\text{max}}$

PALS with quantiles

PALS is not directly suitable for our application:

- f_1 is **not** the expectation of a simulator output,
- we build a GP model for R_{avg} instead.

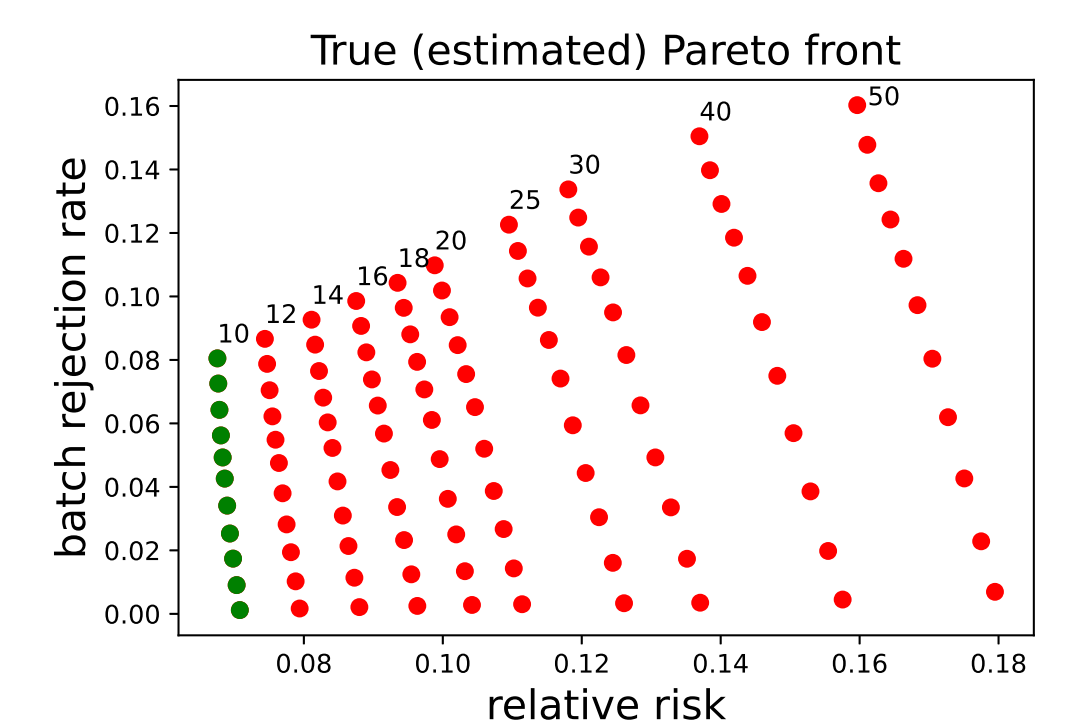
So, we propose using quantiles of QoI to construct confidence rectangles of coverage probability α

$$R_{n,j}^-(x) = q_{n,j}(\alpha/2) \text{ and } R_{n,j}^+(x) = q_{n,j}(1 - \alpha/2)$$

with $q_{n,j}(\cdot)$ being a posterior quantile of $f_j(x)$

Experimental results

- Minimizing f , given $p_{\text{test}}^{\text{milk}} = 1, p_{\text{test}}^{\text{cheese}} = 0.5$, w.r.t.
 - $n_{\text{sample}} \in \{1, 5, 10, 15, \dots, 50\}$
 - $l^{\text{sort}} \in \{10, 12, 14, 16, 18, 20, 25, 30, 40, 50\}$
- **True Pareto front**: estimated using 5000 samples for each of 11×10 input pair
 - A GP regressor is used to smooth the predictions

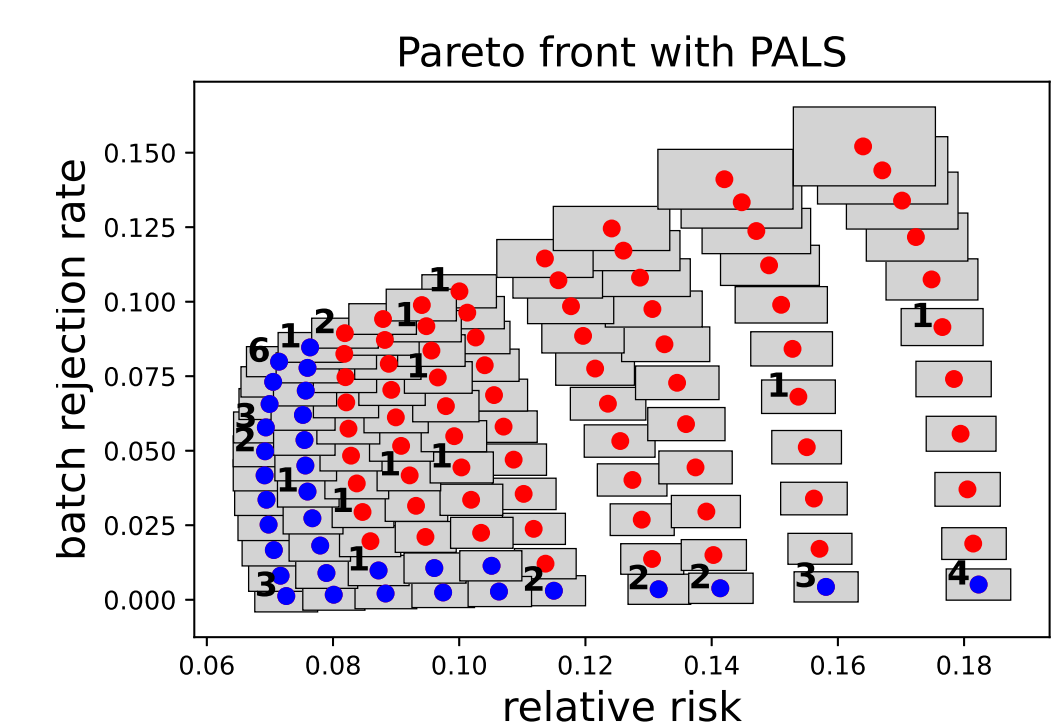


(stripes corresponding diff. values of l^{sort})

– Pareto optimal (green) and dominated points (red)

- **Pareto front estimated using PALS**

– $|I_0| = 10, n_{\text{max}} = 30, \text{samples/evaluation} = 200$



(integers denoting evaluation counts)

- With PALS using significantly less (40×200) budget, the user can provide the following insights
 - Points with $l^{\text{sort}} > 12$ are dominated (red)
 - $l^{\text{sort}} \leq 12$ remain unclassified (blue)

Future work

- Integrate milk loss as objective and $p_{\text{test}}^{\text{milk}}, p_{\text{test}}^{\text{cheese}}$ as design variables
- Take the correlation between outputs into account

References

- F. Perrin, F. Tenenhaus-Aziza, V. Michel, S. Miszczycha, N. Bel, and M. Sanaa. Quantitative risk assessment of haemolytic and uremic syndrome linked to O157:H7 and non-O157:H7 shiga-toxin producing escherichia coli strains in raw milk soft cheeses. *Risk Analysis*, 35(1):109–128, 2014.
- M. Zuluaga, G. Sergeant, A. Krause, and M. Püschel. Active learning for multi-objective optimization. In *Proceedings of the 30th International Conference on Machine Learning*, pages 462–470. PMLR, 2013.
- B. Barracosa, J. Bect, H. Dutrieux Baraffe, J. Morin, J. Fournel, and E. Vazquez. Extension of the Pareto Active Learning method to multi-objective optimization for stochastic simulators. In *SIAM Conference on Computational Science and Engineering (CSE21)*, Mar 2021.

This work is part of the [ArtiSaneFood](#) project (grant number : [ANR-18-PRIM-0015](#)) which is part of the [PRIMA](#) program supported by the [European Union](#).