QUANTITATIVE RISK ASSESSMENT AND OPTIMIZATION OF PROCESS INTERVENTION PARAMETERS FOR FRENCH RAW MILK SOFT CHEESE.

Subhasish Basak[1,2], Julien Bect[2], Laurent Guillier[1], Fanny Tenenhaus-Aziza[3], Janushan Christy[4], Emmanuel Vazquez[2]

[1]Agence nationale de sécurité sanitaire (ANSES), 94700, Maisons-Alfort, France
[2]Université Paris-Saclay, CNRS, CentraleSupélec, L2S, 91190, Gif-sur-yvette, France
[3]Centre national interprofessionnel de l'économie laitière (CNIEL), 75009, Paris, France
[4]Centre technique d'expertise agroalimentaire (ACTALIA), 74800, La-Roche-sur-Foron, France

Context

This study is a part of the European project ArtiSane-Food that aims at controlling food-borne pathogens in artisanal fermented food of meat and dairy, origin produced in the Mediterranean region. As a participating country one of the main objective for France is to allow the continuation of the production of raw milk soft cheeses, which is today potentially at risk due to future European regulations. At the French national level this project is in collaboration with ANSES, CNIEL, CentraleSupélec - Université Paris-Saclay and organizations from the dairy industry.

Problem Statement

The primary goal is to establish efficient intervention strategies, 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.

Intervention strategies in cheese making:

- Pre-harvest milk sorting:
  STEC and *E. coli* strains follow same fecal route!
  - A bulk tank of milk is tested with probability $p^{\text{milk}}_{\text{test}}$
  - Farms with *E. coli* conc. $> l^{\text{sort}}$ are rejected
  - C_{pre}: Cost of testing and rejecting bulk tank milk
- Post-harvest cheese sampling:
  - A batch of cheese is tested with probability $p^{\text{cheese}}_{\text{test}}$
  - From a single batch n_{sample} cheeses are tested for presence of STEC
  - C_{post}: Cost of testing and rejecting cheese batches

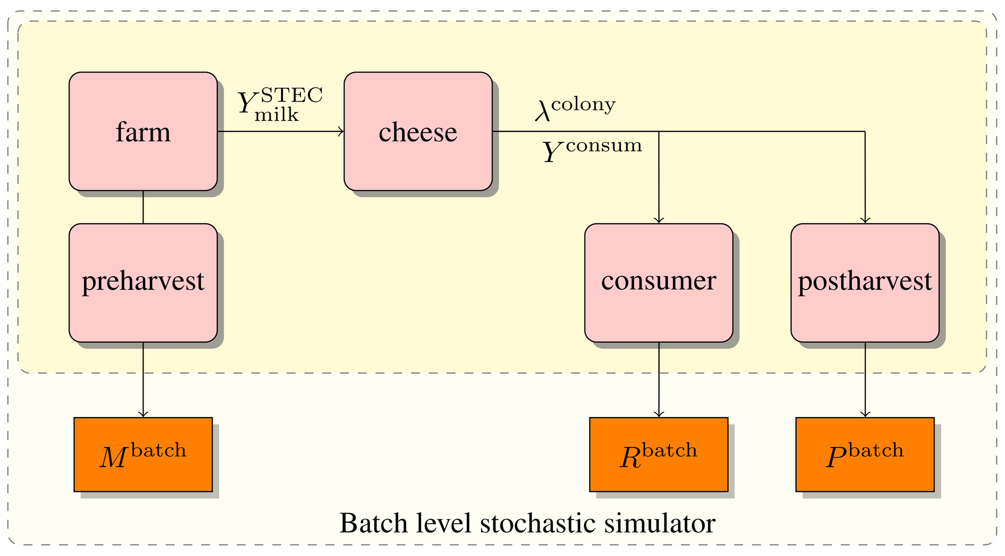
The aim is to find the optimal values of the process intervention parameters $\{p^{\text{milk}}_{\text{test}}, l^{\text{sort}}, p^{\text{cheese}}_{\text{test}}, n_{\text{sample}}\}$, that minimize the risk of HUS and the costs (C_{pre} and C_{post}).

Quantitative Risk Assessment (QRA)

QRA based on model proposed by Perrin et al. (2014)

- Farm module + Pre-harvest step
  STEC conc. $Y^{\text{STEC}}_{\text{milk}}$ in farm milk is computed
  $$Y^{\text{STEC}}_{\text{milk}} = Y^{\text{EC}}_{\text{milk}} \cdot (Y^{\text{STEC}}_{\text{feces}} / Y^{\text{EC}}_{\text{feces}})$$
- Cheese module
  Evolution of STEC is modeled with ODEs
  $$\frac{\mathrm{d}y}{\mathrm{d}t} = \mu^{\max}(t) \cdot y(t) \cdot \left(1 - \frac{y(t)}{y^{\max}}\right)$$
  STEC cells form colonies (clusters) inside cheese
  - No. of colonies (Poisson): N^{colony}
  - Size of colonies (LogNormal): Y^{colony}
- Consumer module
  Batch risk is computed using a dose-response model:
  $$\Gamma = \sum_s N^{\text{colony}}_s \cdot Y^{\text{colony}}_s$$
  $$R^{\text{batch}} = \sum_{\text{age}} g_{\text{age}} \int_\Gamma P[\text{HUS}|\gamma, \text{age}] \cdot \text{p}(\gamma)\mathrm{d}\gamma$$
  averaging over consumer age
- Post-harvest module
  Proportion of rejected batches P^{batch} is computed
  $$P^{\text{batch}} = P[\Gamma > 0] = 1 - \exp(-K \cdot n_{\text{sample}})$$

Quantities of Interest (QoI)



Several batches are simulated to estimate

- $R_{\text{avg}} = \mathbb{E}[R^{\text{batch}} \cdot (1 - P^{\text{batch}} \cdot p^{\text{cheese}}_{\text{test}})]$
- $P_{\text{avg}} = \mathbb{E}[P^{\text{batch}} \cdot p^{\text{cheese}}_{\text{test}}]$
- $M_{\text{avg}} = \mathbb{E}[M^{\text{batch}}]$

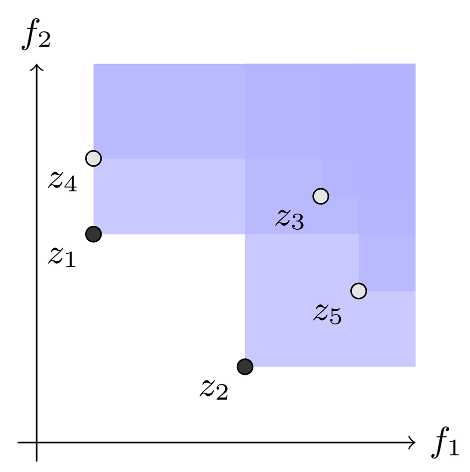
QoIs are

- Relative risk: $f_1 = \frac{R_{\text{avg}}}{(1-P_{\text{avg}})\cdot K_1}$, ($K_1$: baseline risk)
- Intervention cost: $f_2 = C_{\text{pre}}(M_{\text{avg}}) + C_{\text{post}}(P_{\text{avg}})$

Bi-objective optimization

We consider the bi-objective optimization problem
$$\min_{x \in \mathbb{X}} f(x)$$
where, $f = (f_1, f_2)$

- Not necessarily has a unique solution $x^{\text{opt}} \in \mathbb{X}$, in presence of conflicting objectives
- The solution set $\mathcal{P}$ consists of 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 stochastic setting, we assume additive noise: for $x_i \in \mathbb{X}$, we observe $Z_i = f(x_i) + \varepsilon_i$, $\varepsilon_i \sim \mathcal{N}(0, \Sigma)$
- The problem boils down to estimating $\mathcal{P}$

PALS

Optimization of the QRA simulator

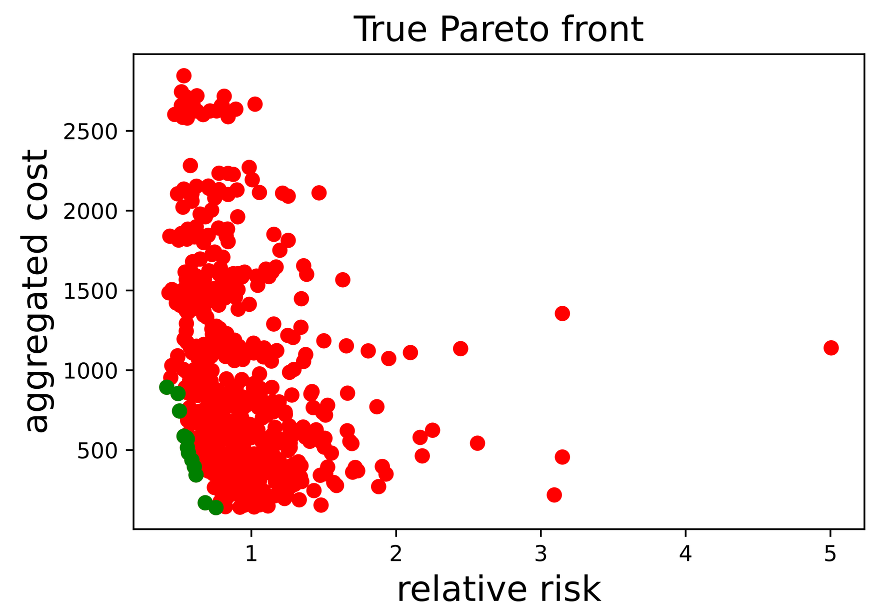
- It is stochastic and computationally expensive
- Gradient based optimization is not feasible
- Thus we reside on Bayesian approaches

Pareto Active Learning for Stochastic simulators proposed by Zuluaga et al. (2013) and extended by Barracosa et al. (2021).

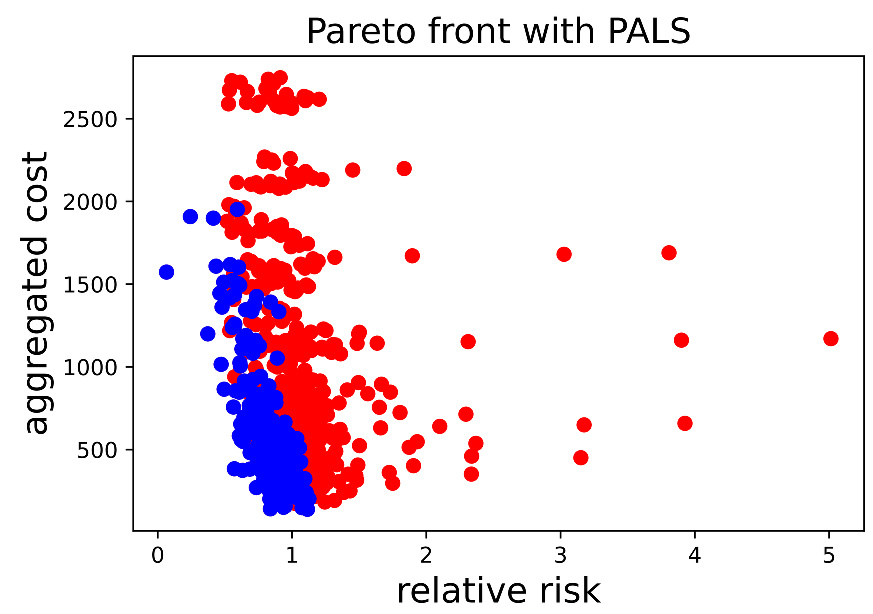
- It uses Gaussian process regression for approximating the simulator function
- Estimates $\mathcal{P}$ by classifying each point in $\mathbb{X}$ as Pareto optimal, Non-Pareto optimal and Unclassified

Experimental results

- Minimizing f over the input space $\mathbb{X}$
  - $n_{\text{sample}} \in \{5, 10, 20, 30, 50\}$
  - $l^{\text{sort}} \in \{10, 20, 30, 40, 50\}$
  - $p^{\text{milk}}_{\text{test}} \in \{1/10, 1/20, 1/30, 1/40, 1/50\}$
  - $p^{\text{cheese}}_{\text{test}} \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$
- True Pareto front: estimated using 5000 samples for each of $5 \times 5 \times 5 \times 5 = 625$ input points



  - Pareto optimal (green) and dominated points (red)
- Pareto front estimated using PALS
  - Initial design size $= 60$, evaluation budget $= 40$
  - batch size per iteration $= 300$



- With PALS using significantly less (100×300) evaluations, the user can provide the following insights
  - Most of the dominated (red) points are well classified
  - The points corresponding to $\mathcal{P}$ remain unclassified (blue)

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. Sergent, 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.