ArtiSaneFood Annual Meeting

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

Bologna, Italy, 7-8 July, 2022

[1]Agence Nationale Sécurité Sanitaire Alimentaire Nationale (ANSES), Maisons-Alfort, France
[2]Université Paris-Saclay, CNRS, CentraleSupeléc, L2S, Gif-sur-yvette, France
[3]Centre National Interprofessionnel de l'Economie Laitière (CNIEL), Paris, France
[4]Centre technique d'expertise agroalimentaire (ACTALIA), La-roche-sur-foron, France

Context of my thesis

- ArtisaneFood France: Study impact of intervention strategies (WP 7)
  - Product: Camembert de Normandie (fromage au lait cru)
  - Pathogen: Shiga Toxin producing Escherichia coli (STEC)
  - Disease: Haemolytic Uremic Syndrome (HUS)
- Thesis duration: January 2021 - December 2023
- Academic advisors:
  - Julien BECT (L2S) & Emmanuel VAZQUEZ (L2S)
  - Laurent GUILLIER (ANSES)
- Industrial advisors:
  - Fanny TENENHAUS AZIZA (CNIEL)

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.

Contents

1 Building QRA simulator

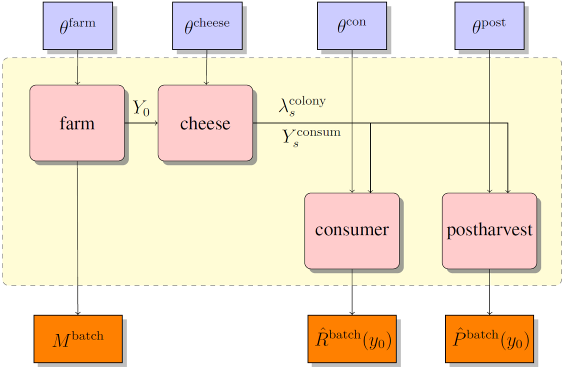
2 Publishing the QRA model

3 Study impact of intervention strategies

4 Work in progress!

1 Building QRA simulator

- QRA model proposed by Perrin et al. (2014)



Batch level simulator

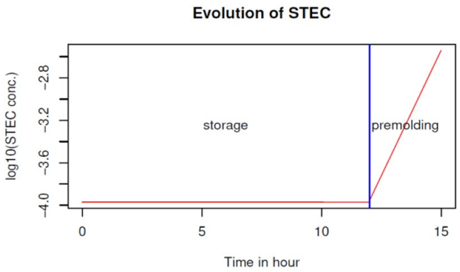
- Implemented in **R**
- Models farm-to-fork continnum
- Simulates fabrication of one cheese batch
  - Contains $22,000$ to $23,000$ cheese (250 g)
- Outputs:
  - Milk loss due to farm rejection
  - Risk of HUS from MPS-STEC
  - Probability of batch rejection

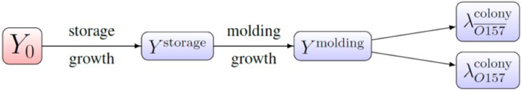
The modules

- Farm module
  - Collection of milk from farms
  - Pre-harvest intervention (milk testing): Test E.Coli conc.
  - Outputs: STEC concentration (CFU/ml) in milk to be used
- Cheese module
  - Evolution of STEC using ODEs
  - STEC cells form clusters (a.k.a colonies) in cheese
  - Outputs
    - Number of colonies in CFU (Poisson)
    - Size of colonies (LogNormal)

Evolution of STEC (liquid phase)

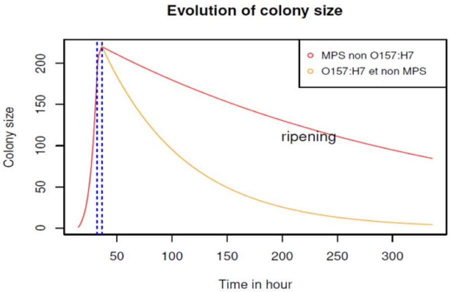
$$\frac{dy}{dt} = \mu^{\max}(t) \cdot y(t) \cdot \left(1 - \frac{y(t)}{y^{\max}}\right)$$

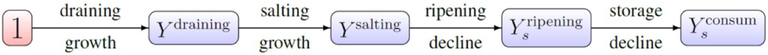




Evolution of STEC (solid phase)

$$Y_s^{\text{consum}} = Y^{\text{salting}} \cdot 10^{-\rho_s \cdot (t^{\text{consum}} \times 24 - t)/24}$$



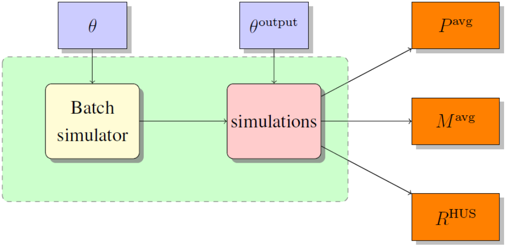


The modules

- Consumer module
  - Risk is computed using a dose response model
    $\Gamma = \sum_s N_{\text{s,sample}}^{\text{colony}} \cdot Y_s^{\text{colony}}$
  - Averaged over consumption behaviour for different ages
  - Outputs Risk of HUS from MPS-STEC
- Post-harvest module
  - Cheese batches are tested for STEC contamination
  - Outputs
    - Probability of rejecting a batch

Final quantitites of interest

- Several batches are simulated



  - Overall risk of HUS
  - Average milk loss
  - Proportion of rejected batches

Modification & improvements

- Bayesian approach to estimate E.Coli distribution parameters
  - Using E.Coli concentration data from CNIEL
- Adaptive algorithm to reduce the computational time
  - Publication : S. Basak, J. Bect , and E. Vazquez. Integration of bounded monotone functions: Revisiting the nonsequential case, with a focus on unbiased Monte Carlo (randomized) methods. In *53èmes Journées de Statistique de la SFdS, Lyon, France, June 2022*
- Analytical solution for estimating batch rejection probabilities
  - Replacing Negative Binomial distributional assumption on dose
- ...

2 Publishing the QRA model

- Food and Ecological Systems Modelling Journal FESMJ
  - Article with description/functionalities of the model (WIP!)
- Food Safety Knowledge Exchange FSKX
  - A FSKX version of the model is available
- RAKIP repository
  - The model will be available on this repository

3 Study impact of intervention strategies

- The QRA simulator is stochastic + expensive
- Pre-harvest intervention (milk testing) parameters
  - Frequency of milk testing
  - Test threshold of E.Coli in CFU/ml
- Post-harvest intervention (cheese testing) parameters
  - Proportion of batches tested
  - Number of cheese samples tested
- Objectives to minimize
  - Risk of HUS
  - Cost due to rejecting milk & cheese batches

Multiobjective optimization

- We consider a biobjective optimization problem

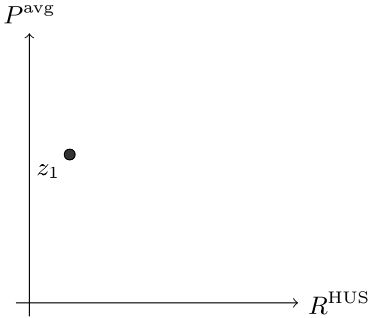
$$\min_{x \in \mathbb{X} \subset \mathbb{R}^2} f(x) \tag{1}$$

- The solution set does not contain a unique optimal solution
- It contains Pareto optimal points

$$\mathcal{P} = \{x \in \mathbb{X} : \nexists x' \in \mathbb{X}, f(x') \prec f(x)\} \tag{2}$$

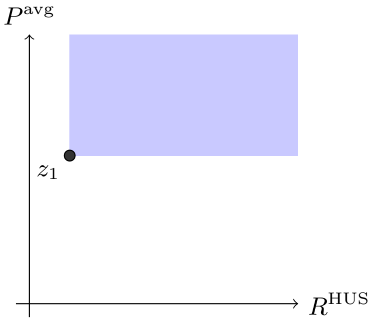
- Where $f' \prec f \implies f'_i \leq f_i, \forall i$, with at least one of the inequalities being strict

Pareto optimal solutions: the objective space



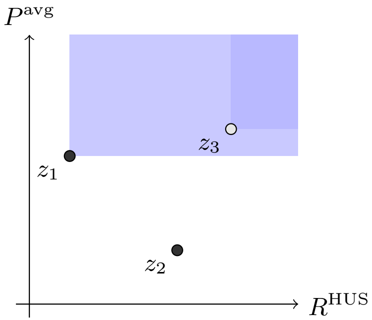
One observation $z_1 = (R_1^{\mathrm{HUS}}, P_1^{\mathrm{avg}})$

Pareto optimal solutions



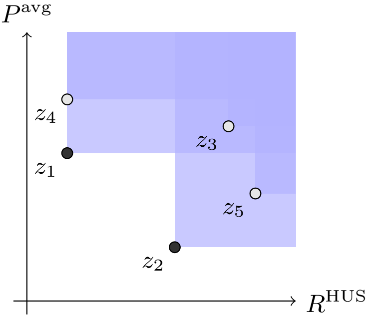
Dominated area by z_1 in objective space

Pareto optimal solutions



z_1 dominates z_3 but not z_2

An illustration: Pareto optimal solutions



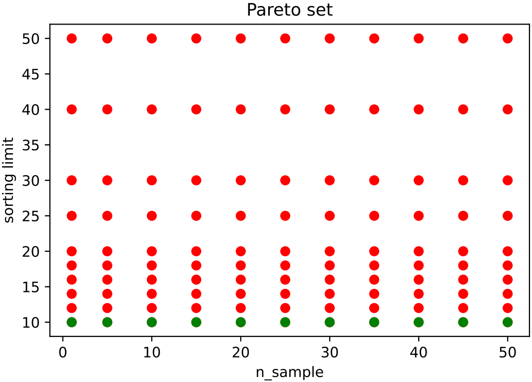
z_3, z_4 & z_5 dominated by z_1 and z_2

The optimization problem: A first approach

- Minimize the function $f = (R^{\mathrm{HUS}}, P^{\mathrm{avg}})$
- To find: optimal value of $(l^{\mathrm{sort}}, n_{\mathrm{sample}}) \in \mathbb{X}$ that minimize f
  - Experimental design $(l^{\mathrm{sort}}, n_{\mathrm{sample}}) \in \mathbb{X}$
    $\mathbb{X} = \{1, 5, 10, 15, \cdots, 50\} \times \{10, 12, 14, 16, 18, 20, 25, 30, 40, 50\}$
  - Objective space $[0, 1]^2$
- A naive solution: Evaluate the simulator $\forall x \in \mathbb{X}$
  - The simulator is expensive
  - The simulator is stochastic
- Therefore, we rely on Bayesian optimization algorithms

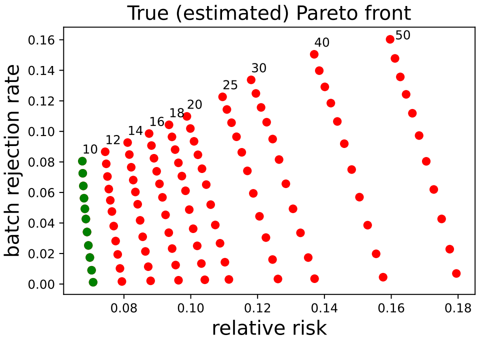
True Pareto set (Experimental design)

- Pareto optimal (green) and non Pareto optimal (red)



True Pareto front (Objective space)

- A heavy Monte-Carlo on $\mathbb{X}$ $(110$ points take 9 hours!)
- Numbers denoting sorting limit l^{sort}

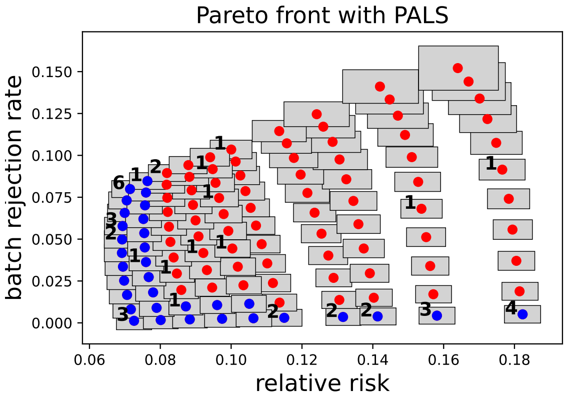


Stochastic Pareto Active Learning (PALS)

- Easy to implement and inexpensive
- Proposed by Zuluaga et al. (2013) and extended by Barracosa et al. (2021)
- Gaussian process (GP) regression is used for surrogate modelling
  - Input space is classified by confidence rectangles using GP posterior mean and variance
- Poster : Subhasish Basak, Julien Bect , Laurent Guillier , Fanny Tenenhaus Aziza, Janushan Christy, Emmanuel Vazquez. Bayesian multiobjective optimization for quantitative risk assessment in microbiology. In *PhD students day in the Annual meeting of GdR MASCOT-NUM research Network, June 2022, Clermont Ferrand, France.*

Estimated Pareto front with PALS

- Pareto optimal (green), unclassified (blue) and non Pareto optimal (red)



4 Work in progress!

- Extending PALS for
  - Grid-free continuous input space
  - more generalized confidence region
- Consider the correlated noise between QoIs
- Consider other objectives like cost of milk loss, batch loss
- Integrate other design variables
  - Proportion of milk batch tested
  - Proportion of cheese batch tested

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, volume 28 of Proceedings of Machine Learning Research, pages 462–470, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.

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), Virtual Conference originally scheduled in Fort Worth, Texas, United States, Mar 2021.