

Multipathogen quantitative risk assessment in raw milk soft cheese, monotone integration and Bayesian optimization

Subhasish Basak^{1,2}

PhD defense

March 20, 2024 - Paris-Saclay, France

Supervised by

Julien Bect², Laurent Guillier¹, Fanny Tenenhaus-Aziza³ & Emmanuel Vazquez²

1. Agence Nationale de Sécurité Sanitaire (ANSES), Maison-Alfort, France
2. Université Paris-Saclay, CNRS, CentraleSupélec, L2S, Gif-sur-Yvette, France
3. Centre national interprofessionnel de l'économie laitière (CNIEL), Paris, France



Innovative Bio-interventions and Risk Modelling Approaches for Ensuring Microbial Safety and Quality of Mediterranean Artisanal Fermented Foods



ArtiSaneFood France

- Potential pathogen contamination - Raw milk soft cheese
“fromages au lait cru” - milk heated < 40°C
 1. Shiga toxin-producing *Escherichia coli* (STEC)
 2. *Salmonella*
 3. *Listeria monocytogenes*





RÈGLEMENT (CE) N° 2073/2005 DE LA COMMISSION
du 15 novembre 2005
concernant les critères microbiologiques applicables aux denrées alimentaires
(Texte présentant de l'intérêt pour l'EEE)
(JO L 338 du 22.12.2005, p. 1)



Fromages au lait cru

Risques : Listeria monocytogenes (agent responsable de la listérose)
Motif : présence de Listeria Monocytogenes (...)

Lait et produits laitiers 25/07/2023

■ Intervention practices - Food safety regulations

Test farm milk

→ everyday ? every 10 days ?

→ *E. coli* contamination level ?

Test cheese batches

→ Test every batch ? (expensive!)

→ How many samples to test ?

Context

- Goal: Ensure the safety of raw milk soft cheese!
 - Optimize the interventions for 3 pathogens simultaneously
- Quantities to ↓ minimize
 - Risk for consumer
 - trade ↑ ↓ off
 - Cost for producer
- Multiobjective optimization - optimal intervention parameters

p_{milk} , p_{cheese} : proportion of farm milk/cheese batches tested

l_{milk} : milk contamination limit n_{sample} : cheese samples tested

Workflow

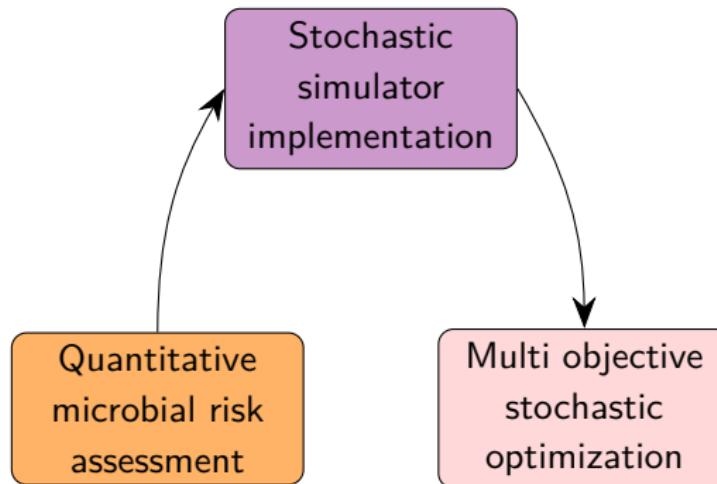


Figure 1: Proposed workflow in the context of this thesis.

Contents

1 Quantitative Microbial Risk Assessment

2 Multiobjective Simulation Optimization

3 Contributions and perspectives

Chapter 1

QMRA - Quantitative Microbial Risk Assessment



1 Quantitative Microbial Risk Assessment

- What is QMRA ? exposure to microbial pathogens → risk of illness
- QMRA literature for raw milk cheese
 1. Pathogenic (MPS) STEC – Perrin et al. (2014), ...
 2. *Salmonella* – Fares (2007), Teunis et al. (2010), ...
 3. *Listeria mono.* – Ricci et al. (2018), Sanaa et al. (2004), ...
- ArtiSaneFood objectives
 - Propose the first multipathogen QMRA model for raw milk cheese
 - Use QMRA model to optimize intervention steps

How to build a QMRA model?

- Framework established by Codex Alimentarius Commission (1999)

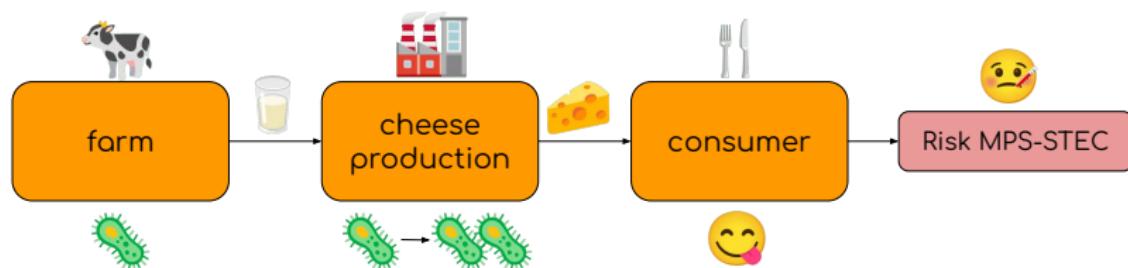


Figure 2: Raw milk cheese Farm-to-Fork QMRA model by Perrin et al. (2014)

Extension to multipathogen framework

- Production of one batch of cheese - 20,000 – 25,000 cheeses

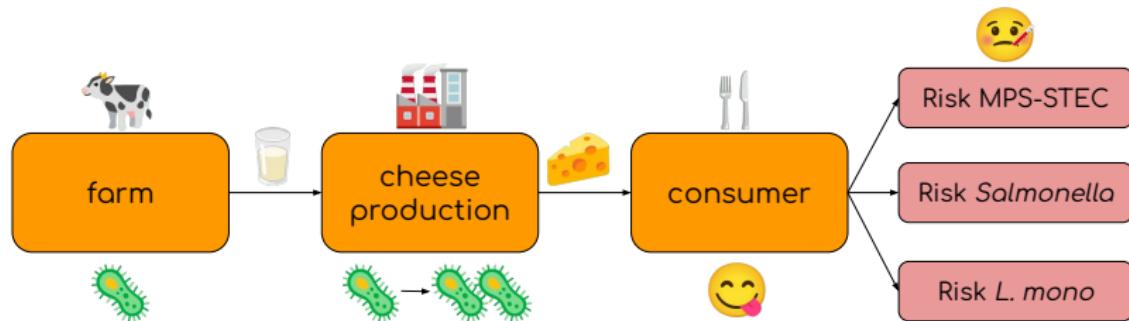


Figure 3: Multipathogen Farm-to-Fork model by Basak et al. (2023b)

Farm module - milk contamination

- Direct approach *Listeria monocytogenes*

$$Y_{\text{Lm}}^{\text{milk}} \sim \text{Lognormal} \text{ (Sanaa et al., 2004)}$$

- Indirect approach STEC and *Salmonella*

$$Y_{\text{pathogen}}^{\text{milk}} = Y_{\text{Ecoli}}^{\text{milk}} \cdot (Y_{\text{pathogen}}^{\text{feces}} / Y_{\text{Ecoli}}^{\text{feces}})$$

→ Why ? Limit of detection due to low concentration

→ How ? Using concentration in animal feces $Y_{\text{pathogen}}^{\text{feces}}$ and $Y_{\text{Ecoli}}^{\text{milk}}$

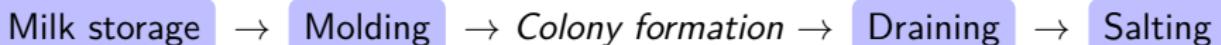
→ Parameters ? Bonifait et al. (2021), Perrin et al. (2014)

- Outputs: Simulated pathogen concentration in milk

Cheese module - pathogen evolution

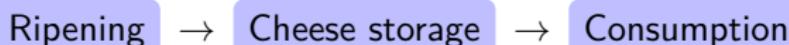
- **Growth ↗ phase**

- Logistic model with rate μ_x (Temp., pH, ...) (Augustin et al., 2005)



- **Decline ↘ phase**

- Decline rates ρ_x (Gonzales-Barron et al., 2022, Perrin et al., 2014)



N.B. **Second growth ↗ phase:** After Salting for *Listeria*

- **Outputs:** Simulated **number** N_x^{colony} and **size** of colonies for **pathogen** x

Consumer module - risk estimation

- Dose in 1 cheese serving $\Gamma_x = (\text{Number} \times \text{size}) \text{ of colonies}_x$
- Risk from 1 cheese serving - Dose-response model

$P_{\text{STEC}}^{\text{illness}}(\text{Age}, \Gamma_{\text{STEC}}) \rightarrow \text{Perrin et al. (2014)}$

$P_{\text{Salmo}}^{\text{illness}}(\Gamma_{\text{Salmo}}) \rightarrow \text{Strickland et al. (2023)}$

$P_{\text{Listeria}}^{\text{illness}}(\text{Age}, \Gamma_{\text{Lm}}) \rightarrow \text{Ricci et al. (2018)}$

- Risk of illness conditional on batch characteristics $\Xi_x = \xi_x$

$$R_x^{\text{batch}}(\xi_x) = \sum_{\text{age}} g(\text{age}) \cdot \mathbb{E}_{\Gamma_x} [P_x^{\text{illness}}(\text{age}, \Gamma_x) \mid \Xi_x = \xi_x]$$

$\rightarrow g(\text{age})$: proportion of cheese consumption per age group

$\rightarrow \Xi_x$: stochastic internal variables simulated by the QMRA model

How to reduce pathogen contamination?

- Intervention steps - food safety regulations

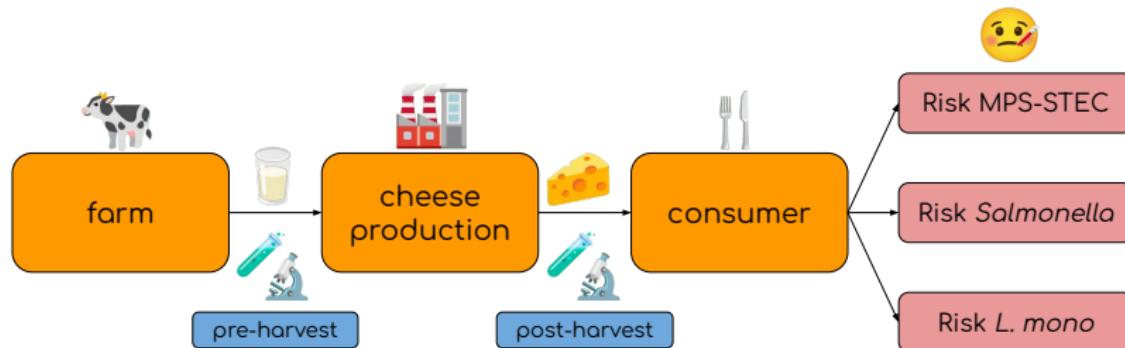


Figure 4: Implementation of **pre-harvest** and **post-harvest** intervention steps

Intervention - reduce contamination

- **Pre-harvest** Testing farm milk with frequency p_{milk}
 - Rejecting farms with high $E. \text{coli}$ contamination in milk
$$Y_{\text{Ecoli}}^{\text{milk}} > l_{\text{milk}} \text{ CFU}$$
- **Post-harvest** Testing cheese batch with frequency p_{cheese}
 - $P[\text{at least one tested samples is contaminated by pathogen } x]$
$$P_x^{\text{reject}}(\Xi_x) = 1 - (1 - P[N_x^{\text{colony}} > 0])^{n_{\text{sample}}}$$
 - Cheese batch is rejected if contaminated by any pathogen x

Outputs of QMRA model

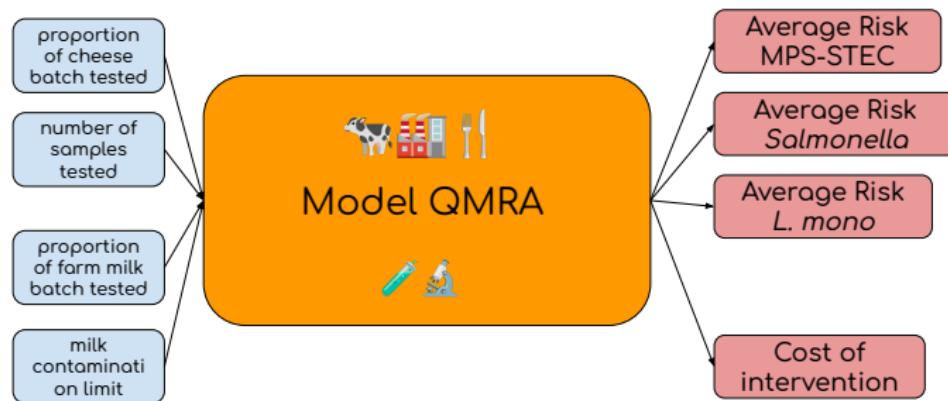
- One random simulated cheese batch conditional on $\Xi = (\Xi_x)_x$
 1. Batch risk $R_x^{\text{batch}}(\Xi_x)$
 2. Amount of Milk rejected M^{batch}
 3. Prob. of rejecting a cheese batch $P^{\text{batch}}(\Xi)$
- Average outputs for a cheese producer
 1. Average risk (pathogen x)

$$R_x = \frac{\mathbb{E}[R_x^{\text{batch}}(\Xi_x) \cdot (1 - P^{\text{batch}}(\Xi) \cdot p_{\text{cheese}})]}{\mathbb{E}[1 - P^{\text{batch}}(\Xi) \cdot p_{\text{cheese}}]}$$

2. Average cost

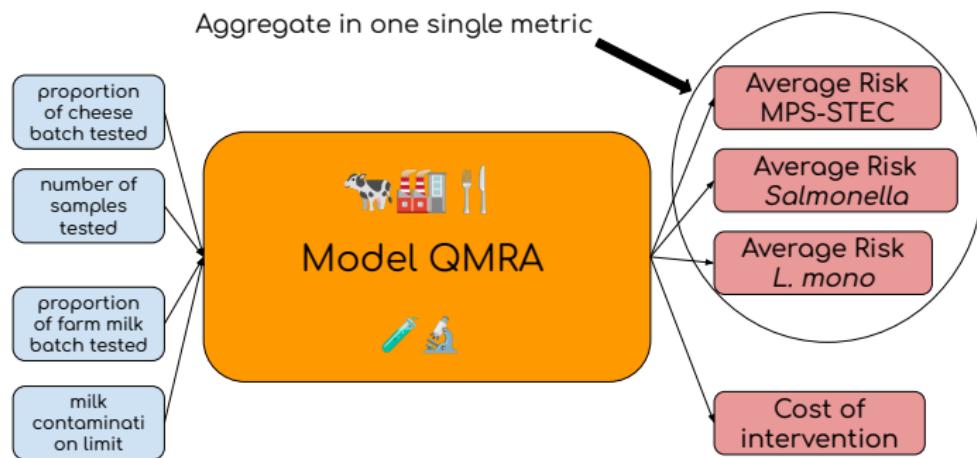
$$C = c_1 + c_2 \cdot \mathbb{E}[M^{\text{batch}}] + c_3 \cdot \mathbb{E}[P^{\text{batch}}(\Xi) \cdot p_{\text{cheese}}], \quad (c_i \text{'s} = \text{cost})$$

Multi-risk QMRA model for 3 pathogens



- **Model validation:** Compare risk/prevalence with previous studies
- **Aim:** Find **optimal inputs** that minimize all the outputs

How to minimize 3 risks simultaneously?



- Different pathogens have **different impact** on public health

DALY: Disability Adjusted Life Years

- DALY = YLL (Years Lost) + YLD (Years with Disability)
- Metric: Average DALY from consuming a cheese portion contaminated by pathogen x from a batch that was not rejected

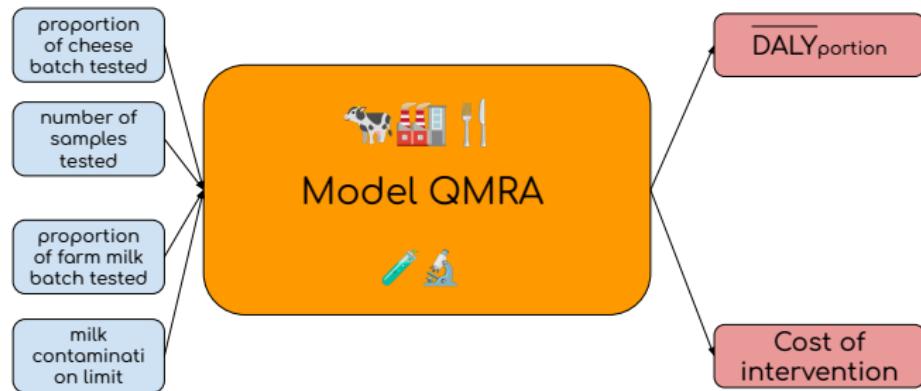
$$\overline{\text{DALY}}_{\text{portion},x} = \sum_{\text{age}} \overline{\text{DALY}}_{\text{illness}(x)}^{\text{portion}}(\text{age}) \cdot R_x(\text{age}) \cdot g(\text{age})$$

- Problem: Lack of epidemiological studies to estimate $\overline{\text{DALY}}_{\text{illness}(x)}^{\text{portion}}(\text{age})$
- Simplified formulation:

$$\overline{\text{DALY}}_{\text{portion},x} = R_x \cdot \text{DALY}(1 \text{ case})_x$$

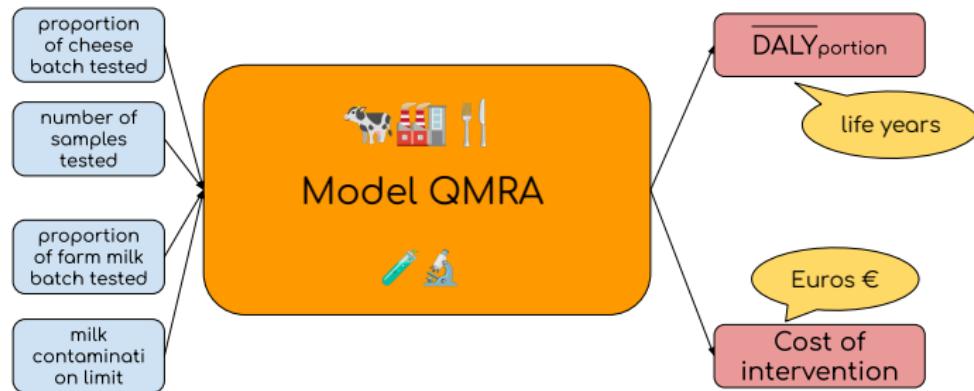
- $\text{DALY}(1 \text{ case})_x$: DALY for 1 case of illness(x) (Cassini et al., 2018)

Multipathogen risks measured in life years



- Combined DALY per portion $\overline{\text{DALY}}_{\text{portion}} = \sum_x \overline{\text{DALY}}_{\text{portion},x}$

Shall we aggregate again ?

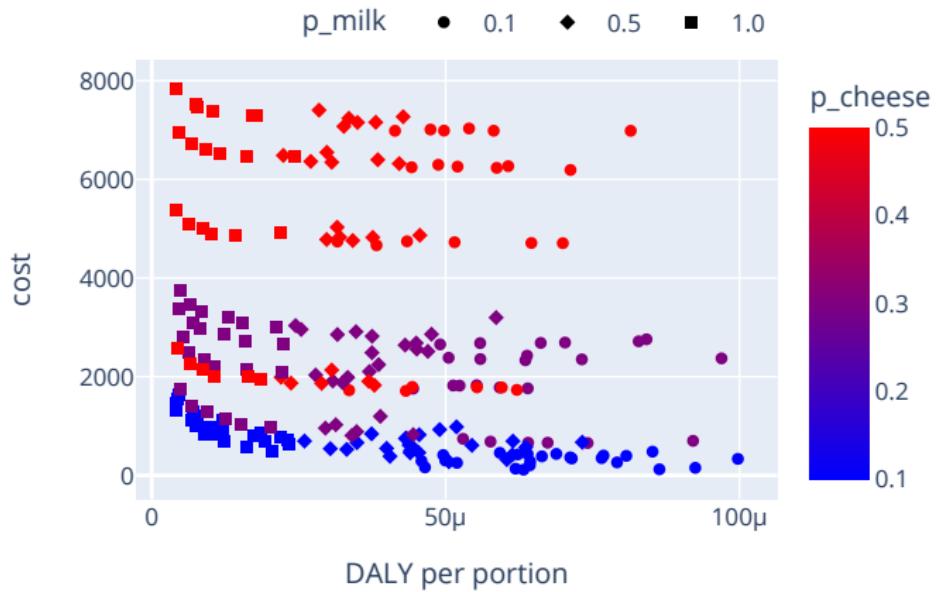


- $\overline{\text{DALY}}_{\text{portion}}$ and cost are measured in different units!

How to choose optimal intervention parameters?

- Consider discrete input space:
 - Prop. of cheese batches tested p_{cheese} : {10%, 30%, 50%}
 - Number of samples n_{sample} : {1, 5, 10, 15}
 - Prop. of milk batches tested p_{milk} : {10%, 50%, 100%}
 - Milk testing limit l_{milk} (CFU/ml): {10, 20, 30, 50, 100, 200}
- Construct a scenario $\rightarrow \{p_{\text{cheese}}, n_{\text{sample}}, p_{\text{milk}}, l_{\text{milk}}\}$
- Total possible scenarios: 216
- 5000 Monte Carlo simulations on each scenario gives DALY & Cost

DALY and intervention cost - 3 pathogens



Is cheese testing really effective?

- *Listeria* has a second growth ↗ phase + unaffected by milk testing
- High prevalence increases batch rejection
- Lower risk of getting *Listeriosis* compared to others pathogen illnesses

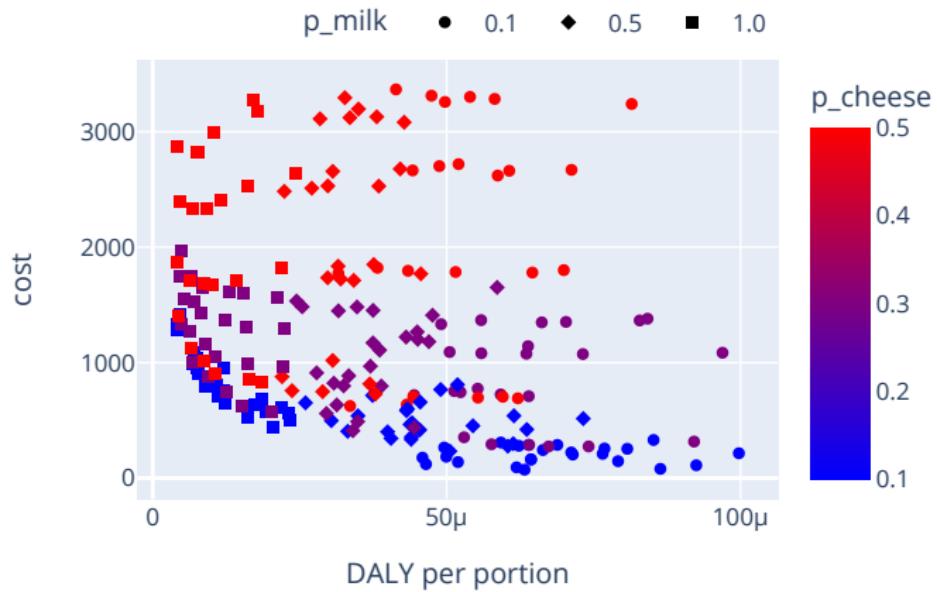
Metric	<i>Listeria</i>	<i>Salmonella</i>	MPS-STEC
Average Prevalence (%)	39.47	0.84	1.97
$R_x/R_{\text{MPS-STEC}}$	4.2×10^{-5}	0.1	1
$\text{DALY}_{\text{portion},x}/\text{DALY}_{\text{portion}}$	4×10^{-5}	5×10^{-4}	0.99
DALY(1 case) _x (Cassini et al., 2018)	3.7	0.019	3.7

Redefine cheese testing plan?

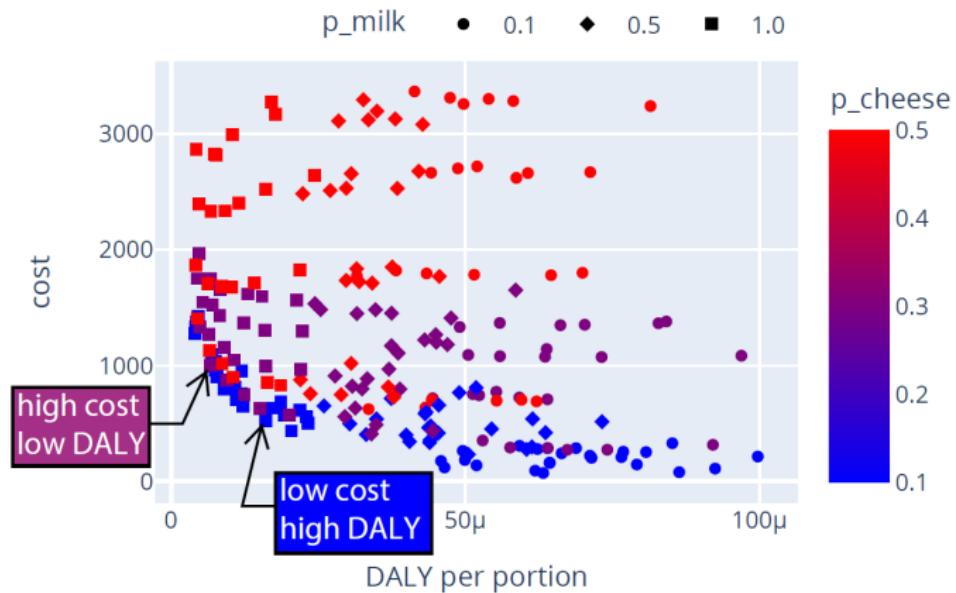
- DALY(1 case)_{Listeriosis} = 3.7 years BUT Listeriosis has very **low risk!**
- Not all *Listeria* strains are highly virulent ! (Pouillot et al., 2024)
- Rejection based on **virulence** / **concentration** threshold?

Metric	<i>Listeria</i>	<i>Salmonella</i>	MPS-STEC
Average Prevalence (%)	39.47	0.84	1.97
$R_x/R_{\text{MPS-STEC}}$	4.2×10^{-5}	0.1	1
$\text{DALY}_{\text{portion},x}/\text{DALY}_{\text{portion}}$	4×10^{-5}	5×10^{-4}	0.99
DALY(1 case) _x (Cassini et al., 2018)	3.7	0.019	3.7

DALY and intervention cost - STEC & Salmonella



DALY and intervention cost - STEC & Salmonella



Key takeaways: chapter 1

- We propose the **first multipathogen QMRA model** (raw milk cheese)
 - Risk assessment (DALY) + Cost estimation (€)
 - Conflicting objectives - **several optimal scenarios**
- Challenges
 - **Noisy** - evaluations - **computationally expensive**
 - $4.5 \text{ seconds} \times 5000 \text{ batches} \times 216 \text{ scenarios} \approx 56 \text{ days} !!$

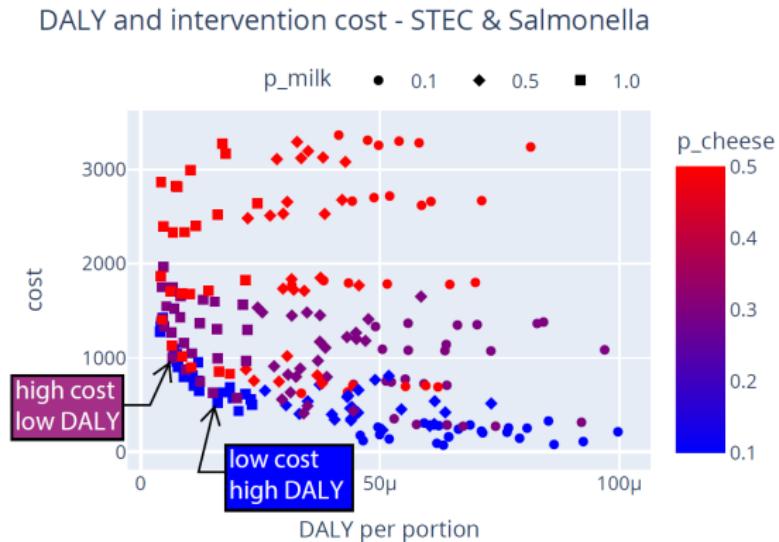
How to find optimal scenarios using limited budget?

Chapter 2

MOSO - Multiobjective simulation optimization

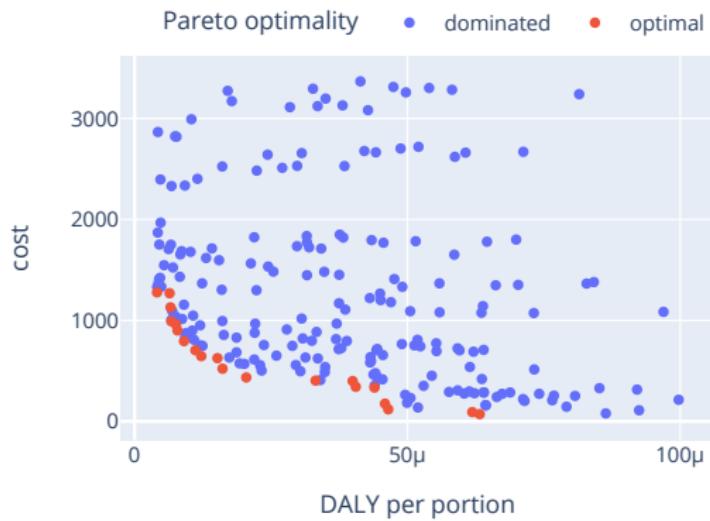


2 Multiobjective Simulation Optimization



Identifying optimal scenarios by evaluating all points

20 optimal cases out of 216 scenarios



First step: problem formulation

- Consider a multiobjective minimization problem of $f = (f_1, f_2, \dots, f_q)$

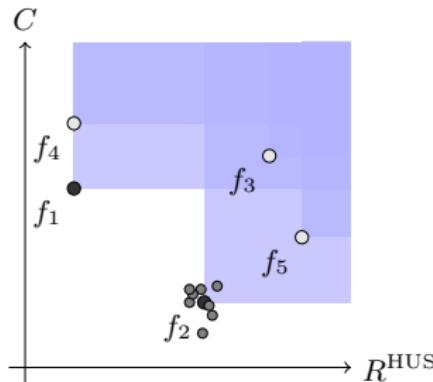
$$\min_{x \in \mathbb{X}} f(x) \quad (1)$$

- Noisy observations: $Z_j^i = f_j(x^i) + \varepsilon_j^i$, with noise $\varepsilon_j^i \sim \mathcal{N}(0, \tau_j^{i^2})$
- Conflicting objectives: No unique optimal solution
- The solution set consists of Pareto optimal points

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

- $f(x') \prec f(x) \implies f_j(x') \leq f_j(x), \forall j$, with at least one strict inequality

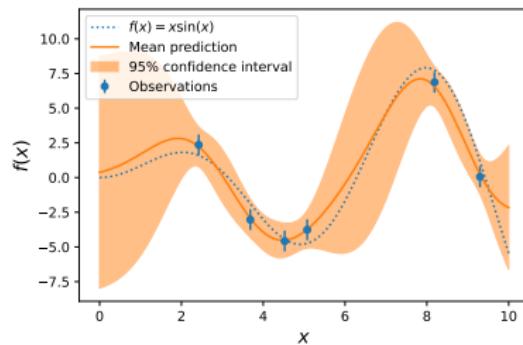
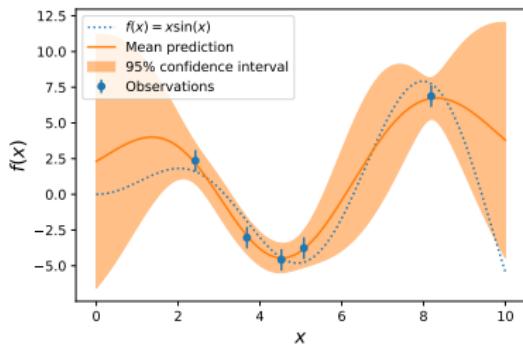
What is Pareto optimality?



f_3, f_4 & f_5 dominated by f_1 and f_2

- Goal: Estimate the Pareto set \mathcal{P} and Pareto front \mathcal{F} (image of \mathcal{P})
- Main challenge → Limited budget + noisy observations

Surrogate models - Gaussian Process (GP) regression



$$Z_j^i = \xi_j(x^i) + \varepsilon_j^i \text{ with } \varepsilon_j^i \stackrel{\text{ind.}}{\sim} \mathcal{N}(0, \tau_j^2), \quad i = 1, 2, \dots \quad (3)$$

- ξ_j s are independent GPs with mean m_j and covariance k_j , $j = 1, \dots, q$
- **Posteriors** of $\xi_j \mid Z_j^i$ are computed by solving system of linear equations

How to use GP regression in optimization?

Algorithm 1 Bayesian Optimization (BO)

Sample* f at n_0 points

while budget > 0 **do**

 Update : GP posterior $\xi \mid Z^1, Z^2, \dots, Z^n$

 Compute : acquisition function $J_n(x)$

 Optimize : $x_{n+1} = \arg \max_{x \in \mathbb{X}} J_n(x)$

 Sample* : f at x_{n+1}

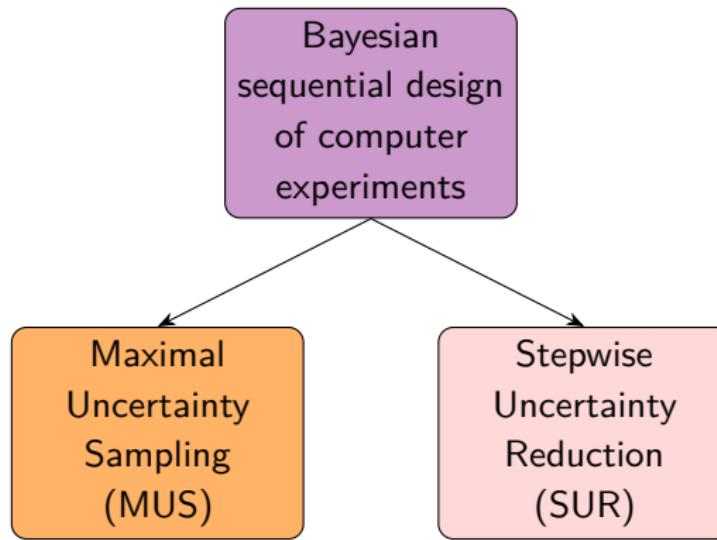
end while

Estimate** $\widehat{\mathcal{P}}$ and $\widehat{\mathcal{F}}$

* f is sampled with a batch size k

** $\widehat{\mathcal{P}}$ and $\widehat{\mathcal{F}}$ are estimated with GP posterior mean $\mu_n(x)$

How to construct an acquisition function $J(x)$?



How to assess performance of a BO algorithm?

1. Volume of Symmetric Difference (VSD) on Pareto front \mathcal{F}

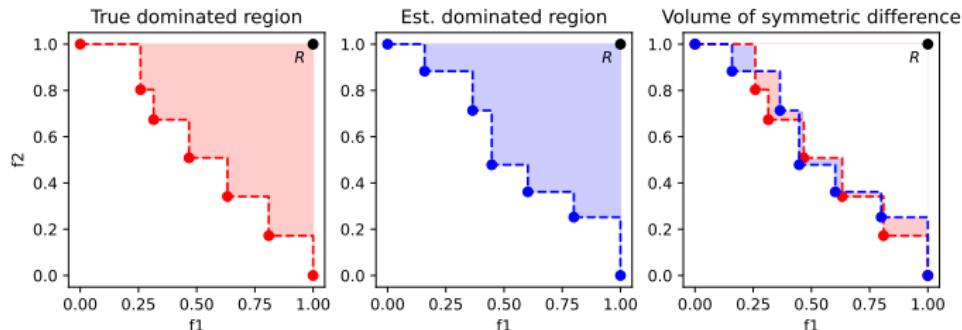
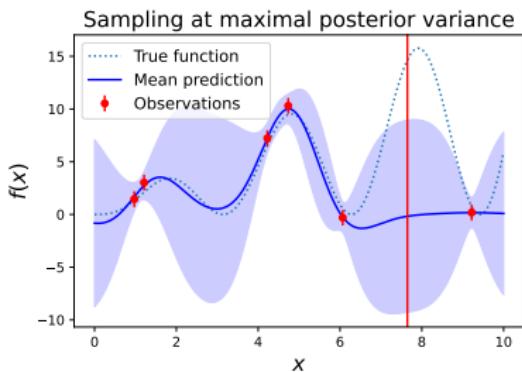


Figure 6: True Pareto front, Estimated Pareto front and reference R

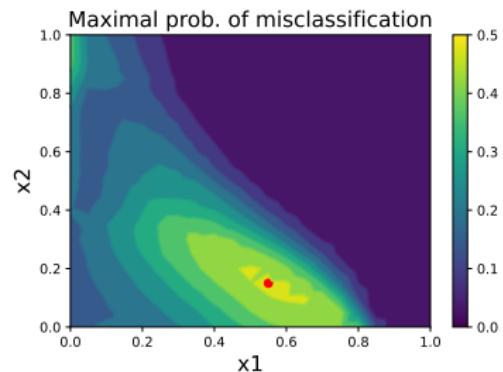
2. Misclassification Rate (MCR) on Pareto set \mathcal{P}

Maximal uncertainty sampling

- Idea : Sample at $x \in \mathbb{X}$ where **uncertainty** is maximum
- Application in different design of experiment frameworks:



(a) Function approximation
Sampling at **maximal posterior variance**



(b) Estimation of probability of failure
Sampling at **maximal misclassification prob.**

A first idea with MUS

- Sampling at maximal probability of misclassification (Bryan et al., 2005)
- $J_n(x)$: Bernoulli variance of the indicator function $\mathbb{1}_{\{x \in \mathcal{P}\}}$

$$\text{Var}_n(\mathbb{1}_{\{x \in \mathcal{P}\}}) = p_n(x) \cdot (1 - p_n(x))$$

→ Estimated with conditional simulations of ξ_n (Binois et al., 2015)

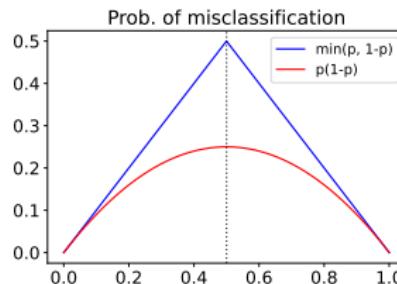


Figure 8: Equivalent measures in literature (Bect et al., 2011)

Performance on test problem g1 from Barracosa et al. (2021)

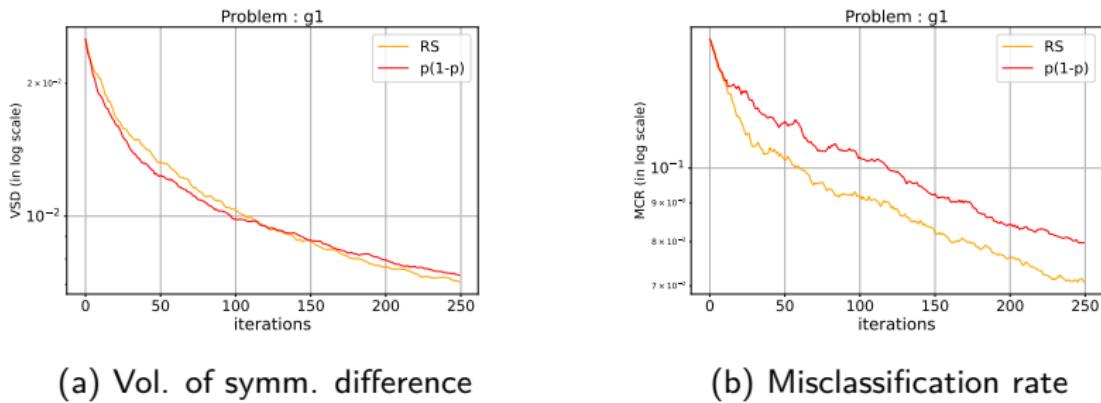


Figure 9: Average performance of $p_n \cdot (1 - p_n)$ based MUS method compared to a naive Random Search (RS) baseline method that samples X_{n+1} randomly from \mathbb{X}

This method does not work!!

Why it fails ?

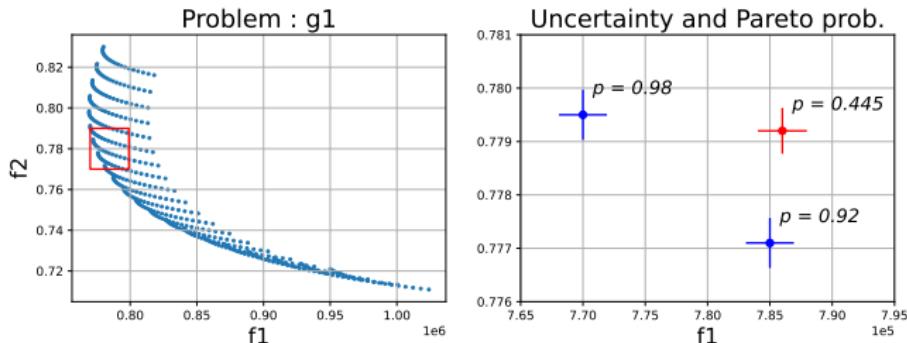


Figure 10: Zoomed from left figure on right, with $p_n(x)$ and post. uncertainty

- Red point $p_n(x) \approx 0.5 \rightarrow$ difficult to classify → due to its neighbors
- The algorithm gets stuck at such points and samples repetitively
- Reducing their post. uncertainty not improving the acquisition criterion

Weighted Mean Squared Error (w-MSE)

- Not all uncertainty measures can be reduced by MUS!
- We try to reduce the weighted mean squared error

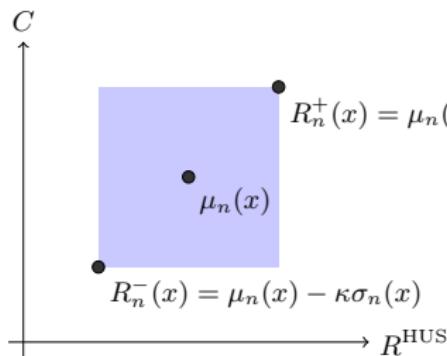
$$X_{n+1} = \arg \max_{x \in \mathbb{X}} \left(w_n(x) \cdot \sum_{j=1}^q \frac{\sigma_{j,n}^2(x)}{R_{j,n}^2} \right)$$

$R_{j,n}$ is a normalizing constant for $j = 1, 2, \dots, q$ -th objective

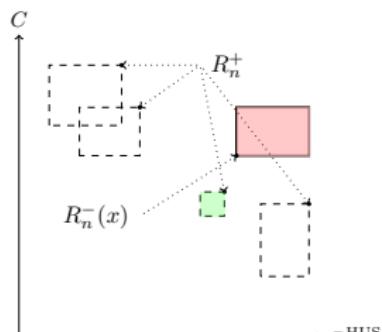
- Weights $w_n(x)$ target the “potential Pareto optimal” region
- Choice of weights ?
 - We proposed $w_n(x)$ based on $p_n(x)$ (see Appendix D)
 - 0 - 1 weights → PAL algorithm (Zuluaga et al., 2013)

PALS (PAL + stochastic setting): Barracosa et al. (2021)

- PALS is a **w-MSE** algorithm with $w_n(x) = \mathbb{1}_{\{x \in \mathbb{X} \setminus N_n\}}$



(a) Confidence rectangle



(b) $R_n^+(x') \prec R_n^-(x)$

- $x \in N_n$ if **optimistic** $R_n^-(x)$ is dominated by some **pessimistic** $R_n^+(x')$

w-MSE methods against PALS (No consistent improvement)

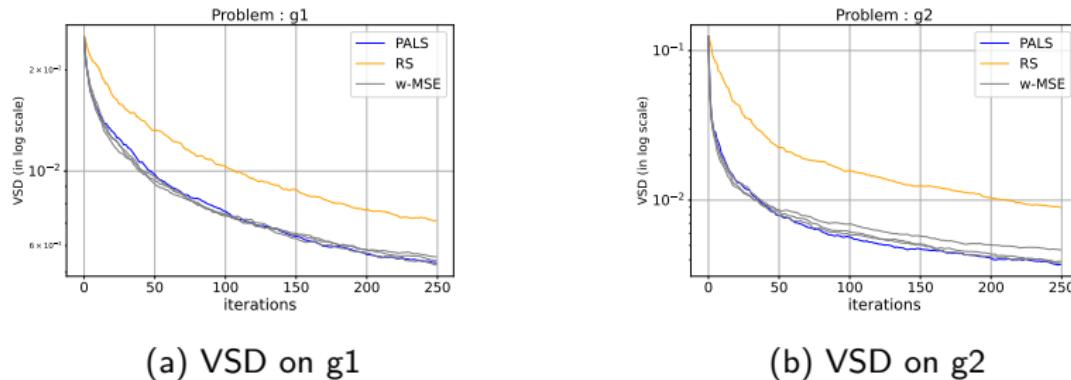


Figure 12: Average performance of PALS with w-MSE methods and Random Search

PALS being a **simple** and **inexpensive** algorithm
 remains difficult to beat in MUS framework!!

Stepwise uncertainty reduction?

- Origins & applications : reliability theory, optimization, ...
 - Vazquez and Martinez (2006), Villemonteix et al. (2007), Vazquez and Bect (2009), ...
- Quantification of uncertainty at step n : H_n
 - Acquisition function is minimized to sample X_{n+1}
$$J_n(x) = \mathbb{E}_n(H_{n+1}|X_{n+1} = x)$$
 - \mathbb{E}_n : conditional expectation w.r.t $\{X_1, Z_1, \dots, X_n, Z_n\}$
 - $J_n(x)$ does not always have a closed analytical form

Weighted Integrated Mean Squared Error (w-IMSE)

- w-IMSE as uncertainty measure H_n gives the following acquisition

$$X_{n+1} = \arg \min_{x \in \mathbb{X}} \sum_{i=1}^{|\mathbb{X}|} w_n(x_i) \sum_{j=1}^q \frac{\sigma_{j,n+1}^2(x_i|x)}{R_{j,n}^2}$$

How to choose a good weight $w_n(x)$ function?

- SUR criterion when $w_n(x_i) = p_n(x_i)$
- Other choice of “plug-in” weights
 - PALS based : $w_n(x_i) = \mathbb{1}_{\{x_i \in \mathbb{X} \setminus N_n\}}$
 - Proposed weights : PALS classification + deviations of \mathcal{F}

Proposed weights - focused on the VSD metric

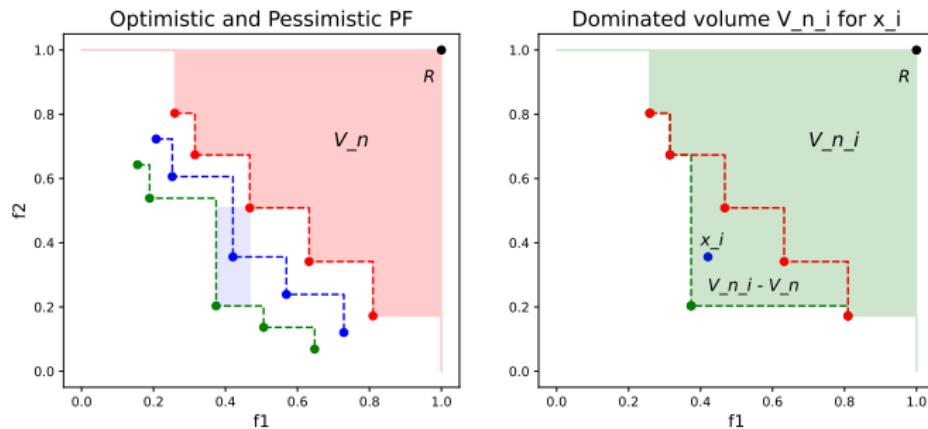


Figure 13: Pareto front $\widehat{\mathcal{F}}$, $\widehat{\mathcal{F}_p}$ and $\widehat{\mathcal{F}_o}$ with dominated volumes V_n and $V_{n,i}$

$$w_n(x_i) \leftarrow \mathbb{1}_{\{x_i \in \mathbb{X} \setminus N_n\}} \cdot (V_{n,i} - V_n)$$

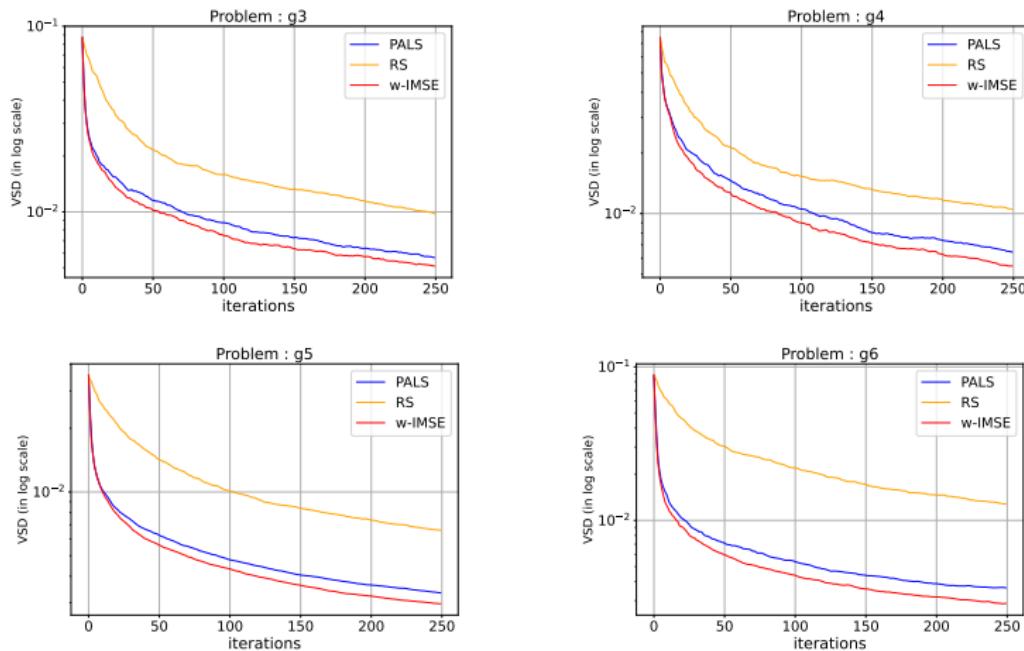


Figure 14: Average VSD metric on problems g-3,4,5,6 (Barracosa et al., 2021)

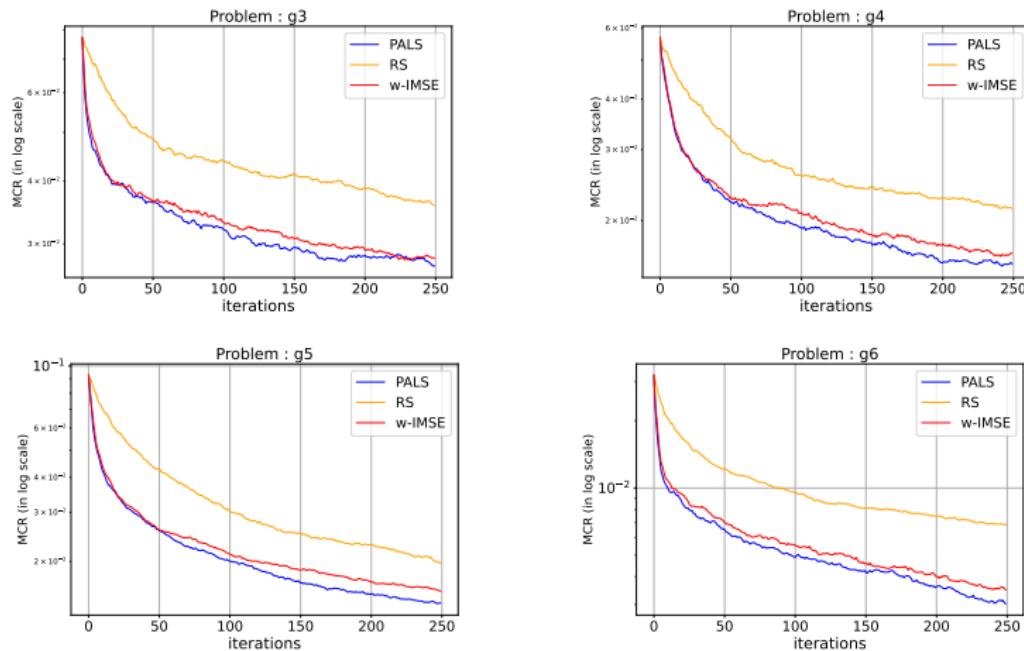


Figure 15: Average MCR metric on problems g-3,4,5,6 (Barracosa et al., 2021)

Key takeaways: chapter 2

- Proposed w-IMSE method **improves consistently over PALS**
 - Improvement the VSD metric on \mathcal{F} (Basak et al., 2023a)
 - Improvement on simulation budget for a given performance
- **However** : Difficult to beat PALS w.r.t both MCR and VSD
- **Remarks on PALS**
 - A **simple, inexpensive, easy-to-implement** MUS algorithm
 - Not easy to beat even with sophisticated SUR methods
- **Application QMRA:** Proposed extension of PALS (Basak et al., 2022a)

Contributions and perspectives



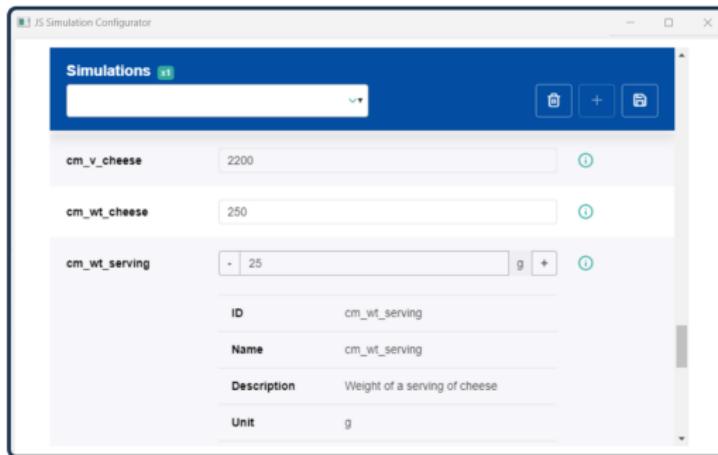
3 Contributions and perspectives

QMRA model development

Ensuring food safety + optimize analytical cost

- Proposed **first multipathogen QMRA model** for raw milk cheese
- ICPMF-12 Japan (Basak et al., 2023b) + Article MRA (under review)
- **Key features:**
 - Assess burden on public health - **DALYs**
 - Estimation of **intervention costs**
 - Parameter estimation with **ArtiSaneFood project data**

Article FESMJ + model FSKX (Basak et al., 2024)



Version 4.5.2
(Build March 23, 2022)



Figure 16: KNIME (open source GPLv3) user interface for FSKX model

Multipathogen model in R-shiny (WIP!)

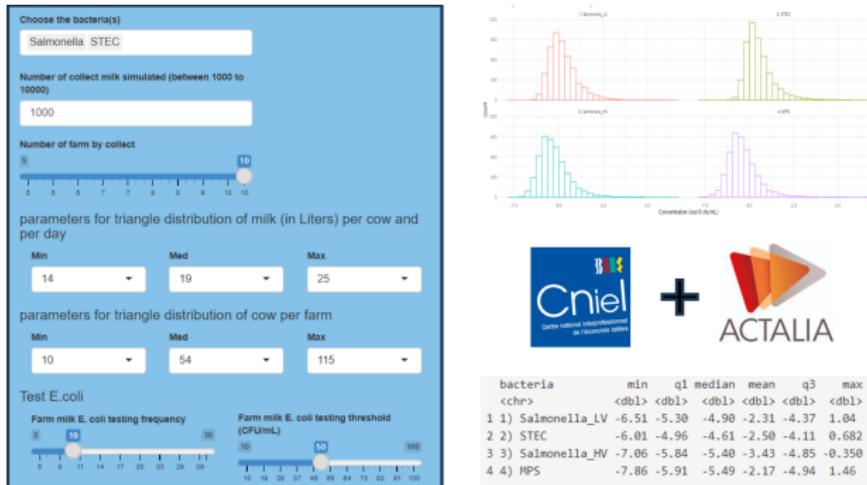


Figure 17: R-shiny application for the **cheese professionals**

QMRA model in optimization

- “Classical” scenario-based analysis is common in QMRA literature
- This becomes impractical with high simulation cost

Novel approach

- Optimization of multipathogen risk and intervention cost simultaneously!
- Approach: Multiobjective simulation optimization (MOSO) algorithms
- Why MOSO algorithms?
 - Optimize expensive stochastic simulator
 - Estimate Pareto optimal solutions in limited budget

Multiobjective simulation optimization

- We study MUS and SUR based MOSO algorithms
- Numerical benchmark against PALS algorithm
- Proposed a w -IMSE based pseudo-SUR algorithm (Basak et al., 2023a)
 - Weights based on the uncertainty on the Pareto front (\mathcal{F})
 - Significant improvement over PALS on estimation of \mathcal{F}
- Remarks on MUS principle based algorithms
 - Well-known sampling criteria from reliability theory literature are inefficient in MOSO framework
 - Not all uncertainty measures can be reduced by MUS!

QMRA perspectives

Model calibration: “*Models are always incomplete representations of the system they are intended to model, but they can still be useful.*” (WHO, 2021)

1. Farm module for *Listeria*
 - Exposure assessment is more difficult than other pathogens
2. Milk testing step depends only on *E. coli*
 - Improve milk testing based on *Listeria* + *Salmonella* detection?
 - Implement inclusion-exclusion of farms + dynamic hygiene parameters
3. Cheese testing for *Listeria*
 - Redefine cheese testing considering virulence / threshold?

MOSO perspectives

1. **QMRA applicability**: objective not an expectation of simulator output
 - Extended PALS (Basak et al., 2022a) algorithm
 - **Perspective**: similar extension for proposed w-IMSE algorithm
 - **Challenge**: Estimation of $\sigma_{n+1}^2(x_i | x)$ inside w-IMSE criterion
 - Use of Monte Carlo simulations from ξ (**expensive**)?
2. **Extension**: **multiobjective** w-IMSE algorithm + **heteroscedastic** noise
3. **Extend numerical benchmark**
 - Ranking and selection (Lee et al., 2010, Rojas Gonzalez et al., 2020)
 - Others (Belakaria et al., 2020, Hernández-Lobato et al., 2014)

Thank you for your attention!



COPIL ArtiSaneFood 2023, Maison du Lait, Paris

A Appendix: QMRA

- **Dose-response models**

Probability of HUS is assumed to follow a binomial process:

$$P(\text{HUS} \mid \text{age}) = 1 - (1 - r_{\text{age}})^{\Gamma}, \text{ with } r_{\text{age}} = r_0 \times \exp(k \times \text{age})$$

k and r_0 estimated from epidemiological surveillance and outbreak data

- **STEC dose-response**

Dose-response model (Perrin et al., 2014) based on epidemiological estimates of the incidence rates of HUS for different population age groups and *E. coli* O157: H7 outbreak data in France.

- **Salmonella dose-response**

Outbreak data for *Salmonella* from food samples have been used to fit dose-response curves (Strickland et al., 2023).

- **Listeria dose-response**

Adapted from Ricci et al. (2018), based on the Poisson model, which takes into account the variability in susceptibility across mutually exclusive population subgroups, as proposed by Pouillot et al. (2015).

- **Model validation**

- Compare QMRA model outputs with previous studies
- Lack of epidemiological studies and relevant data
- Validation of model components with challenge test data

DALY: Disability Adjusted Life Years

- **Aim:** Evaluate the collective impact of all pathogen
- **Metric:** Average DALY from consuming a cheese portion contaminated by pathogen x from a batch that was not rejected

$$\overline{\text{DALY}}_{\text{portion},x} = \sum_{\text{age}} \overline{\text{DALY}}_{\text{illness}(x)}^{\text{portion}}(\text{age}) \cdot R_x(\text{age}) \cdot g(\text{age})$$

For a cheese portion (25g) consumer of Age = age

- Unknown $\overline{\text{DALY}}_{\text{illness}(x)}^{\text{portion}}(\text{age})$: DALY from illness(x) by consuming
- QMRA model output $R_x(\text{age})$: Average risk of illness(x)
- Known $g(\text{age})$: proportion of cheese consumption per age group

DALY: Simplification

- Problem: Lack of epidemiological studies to estimate $\overline{\text{DALY}}_{\text{illness}(x)}^{\text{portion}}(\text{age})$
- Equivalent formulation:

$$\overline{\text{DALY}}_{\text{portion},x} = R_x \cdot \sum_{\text{age}} \overline{\text{DALY}}_{\text{illness}(x)}^{\text{portion}}(\text{age}) \cdot \tilde{g}_\theta(\text{age})$$

– $\tilde{g}_\theta(\text{age})$: Age dist. of cases of illness(x) given QMRA simulator inputs θ

- Simplified formulation:

$$\overline{\text{DALY}}_{\text{portion},x} = R_x \cdot \text{DALY}(1 \text{ case})_x$$

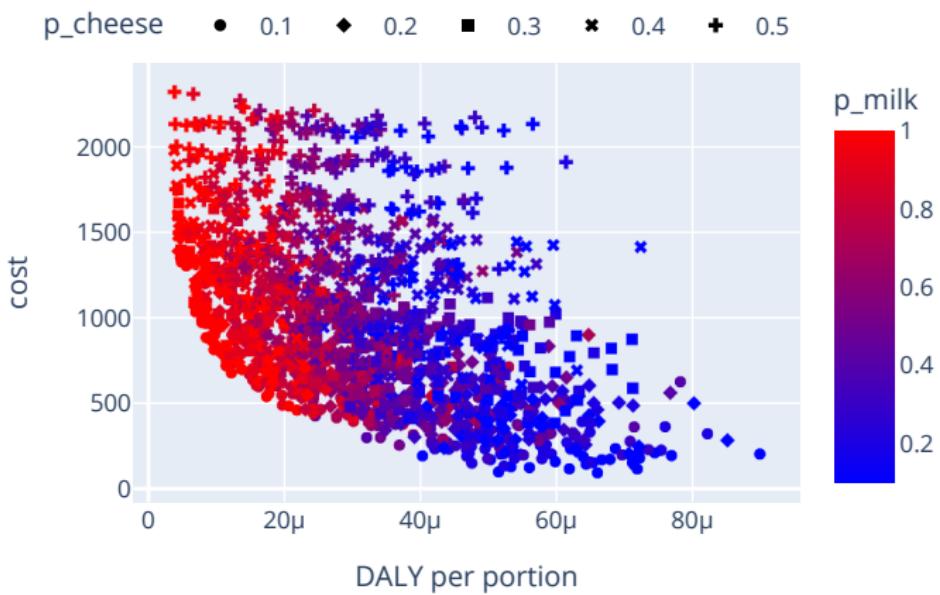
– $\text{DALY}(1 \text{ case})_x$: DALY for 1 case of illness(x) (Cassini et al., 2018)

DALY: Hypotheses & approximations

- (Cassini et al., 2018) estimated DALY(1 case)_x values using EU/EEA epidemiological data
- We want to compute DALYs for cheese consumption in France
- Assumptions:
 - $\tilde{g}_\theta(\text{age}) = P[\text{age} \mid \text{illness}(x)]$ for France closely align to EU/EEA
 - $\tilde{g}_\theta(\text{age})$ remains almost unaffected by QMRA model inputs θ
- Combine DALYs ignoring effects of concurrent instances of illnesses

$$\overline{\text{DALY}}_{\text{portion}} = \sum_x \overline{\text{DALY}}_{\text{portion},x}$$

DALY and intervention cost - STEC



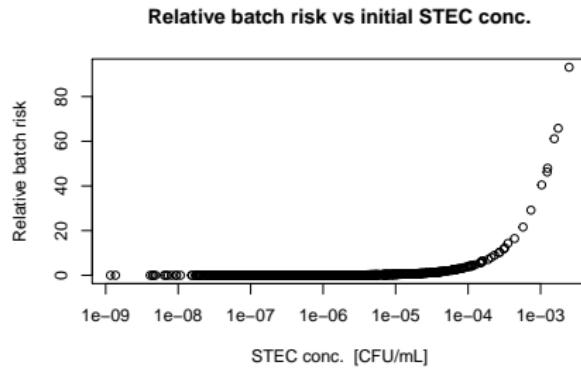
Chapter 2

Integration of monotone bounded functions



B Integration of monotone bounded functions

- Aim: Numerical integration of QMRA **batch level** outputs



- $R_x^{\text{batch}}(Y_x^{\text{milk}})$, $P^{\text{batch}}(Y_x^{\text{milk}})$ are **monotonic** and **bounded** w.r.t Y_x^{milk}

Problem formulation

We are interested in approximating $\mathbb{E}(g(Y)) = \int g(y) P_Y(dy)$

- $g : \mathbb{R} \rightarrow \mathbb{R}$ is **monotonic** and **bounded**
- P_Y is known, CDF F_Y of Y is continuous
- After scaling, the problem reduces to estimating

$$S(f) = \int_0^1 f(x)dx,$$

where $f \in F : [0, 1] \rightarrow [0, 1]$ is a **non-decreasing function**

- **Fixed sample-size setting:** $S(f)$ is estimated from n evaluations of f
- Two categories of methods: **nonsequential** and **sequential** for class F

Nonsequential randomized methods

- Evaluates f at n **random** sample points $\{X_1, X_2, \dots, X_n\}$ in $[0, 1]$
- The estimator of $S(f)$ is constructed as

$$\widehat{S}_n(f) = \varphi(X_1, f(X_1), \dots, X_n, f(X_n))$$

where function $\varphi : [0, 1]^{2n} \rightarrow \mathbb{R}$

- The worst-case L^p error of a method $\widehat{S}_n(f)$ over the class F :

$$e_p(\widehat{S}_n) = \sup_{f \in F} \mathbb{E} \left(\left| S(f) - \widehat{S}_n(f) \right|^p \right)^{1/p} \quad (4)$$

Literature study

- Deterministic methods (Kiefer, 1957):
 - Trapezoidal rule has minimum worst case error $1/2(n + 1)$
- Randomized methods (Novak, 1992):
 - Nonsequential randomized: $e_1(\hat{S}_n) \geq 1/8n$
 - Sequential randomized: $e_1(\hat{S}_n) \geq \sqrt{2}/32 \cdot n^{-3/2}$
- Extending Theorem 1 in Novak (1992), we prove that

Theorem (Basak et al. (2022b))

For any nonsequential randomized method with sample size n ,

$$e_p(\hat{S}_n) \geq (1/2)^{2+1/p} 1/n.$$

Unbiased nonsequential randomized methods

- Simple Monte Carlo estimator $\widehat{S}_n^{\text{MC}}(f) = \frac{1}{n} \sum_{i=1}^n f(X_i)$
- Control variate estimator with $\tilde{f}(X_i) = X_i$

$$\widehat{S}_n^{\text{cv}}(f) = \frac{1}{n} \sum_{i=1}^n (f(X_i) - \tilde{f}(X_i)) + \frac{1}{2}$$

$$\rightarrow X_i \stackrel{iid}{\sim} U_{[0,1]}$$

- Stratified sampling estimator

$$\widehat{S}_n^{\text{str}}(f) = \sum_{k=1}^K w_k \cdot \frac{1}{n_k} \sum_{i=1}^{n_k} f(X_{k,i})$$

$$\rightarrow X_{k,i} \stackrel{iid}{\sim} U_{I_k} \text{ from stratum } I_k = [x_{k-1}, x_k] \text{ with } w_k = \Delta I_k \text{ and } \sum_k n_k = n$$

Maximal squared L_2 error of nonsequential randomized methods

lower bound	$e_2(\widehat{S}_n^{\text{str}})^2$	$e_2(\widehat{S}_n^{\text{cv}})^2$	$e_2(\widehat{S}_n^{\text{MC}})^2$
$1/32n^2$	$1/4n^2$	$1/12n$	$1/4n$

- The best-known variance upper bound
 - $n \leq 2$: Control variate estimator ($\widehat{S}_n^{\text{cv}}$)
 - $n \geq 3$: Stratified sampling estimator ($\widehat{S}_n^{\text{str}}$)
- Deterministic **trapezoidal rule** has maximum squared error $1/4(n + 1)^2$

However

- **Unbiased nonsequential** methods are **building blocks** for **sequential** methods

Sequential randomized methods

- Novak (1992) proposed a **sequential randomized** method based on **stratified sampling** to prove the **optimality** of L_1 error lower bound of order $n^{-3/2}$

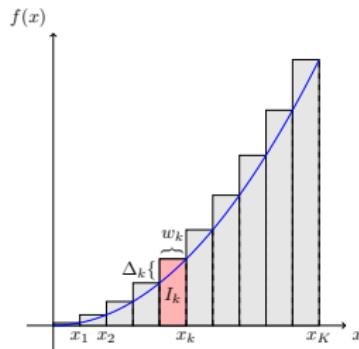


Figure 18: Stratified sampling with equispaced strata

2-step sequential method (Novak, 1992)

- Total budget $n = 3m + 1$ and $(m + 1)$ initial **equispaced** strata I_k
- Budget allocation: based on strata bound difference $\Delta_k = f(x_k) - f(x_{k-1})$

$$n_k = \begin{cases} 1 & \text{if } \Delta_k = 0 \\ \lceil (m + 1) \cdot \Delta_k \rceil & \text{if } \Delta_k > 0, \end{cases} \quad (5)$$

- 2nd step: **Stratified sampling** inside I_k with n_k equispaced substrata
 - lower maximal L_2^2 error of $1/4n_k^2$ compared to $1/4n_k$ of MC
- Novak's unbiased estimator

$$\widehat{S}_n^{\text{novak}}(f) = \sum_{k=1}^{m+1} w_k \cdot \frac{1}{n_k} \sum_{i=1}^{n_k} f(X_{k,i})$$

Improving variance upper bound

- Without modifying the method, we improve the claimed upper bound of variance (Novak, 1992) by a factor 2

Theorem

The two stage sequential randomized algorithm proposed by Novak (1992) satisfies

$$\text{Var}(\widehat{S}_n^{\text{novak}}(f)) \leq \frac{1}{4 \cdot (m+1)^3}.$$

This upper bound is exactly attained by a staircase function that has equal strata bound differences (Δ_k) in each of the strata.

- Improves the L_1 upper bound by a factor $\sqrt{2}$: $(\sqrt{54}/2\sqrt{2}) \cdot n^{-3/2}$

Contributions: Modifications to Novak's method

Aim: Formalizing Novak's method as an algorithm for practical application

1. Saving budget on strata with $\Delta_k = 0$
 - Integral can be **analytically** computed
2. Improved (generalized) budget allocation rule

$$n_k = \lceil \Delta_k \cdot \alpha \cdot (n - m) \rceil$$

where α maximizes $\sum_{k=1}^{m+1} n_k$ subject to budget constraints

3. Choosing optimal initial budget for strata
 - Using $1/6$ of total budget for initial strata splits
 - improves the L_1 upper bound by another factor $\sqrt{2}$: $(\sqrt{54}/4) \cdot n^{-3/2}$

Numerical benchmark

- Two benchmarks: using **staircase** and **smooth** functions
- Compare the **Empirical MSE** and variance **upper bound**, with varying budget
- Baseline integration methods
 1. Trapezoidal Rule
 2. Simple Monte Carlo
 3. Reference method (Novak, 1992)
- Proposed integration methods
 4. **full_alloc** : Reference + improved budget allocation
 5. **full_alloc_init** : **full_alloc** + optimal initial budget

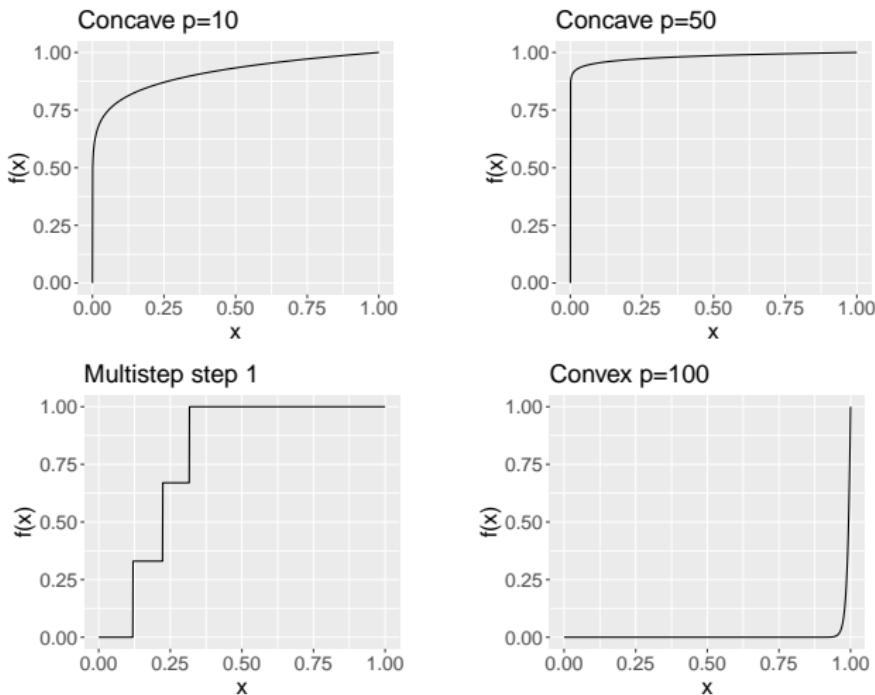


Figure 19: Smooth and step functions

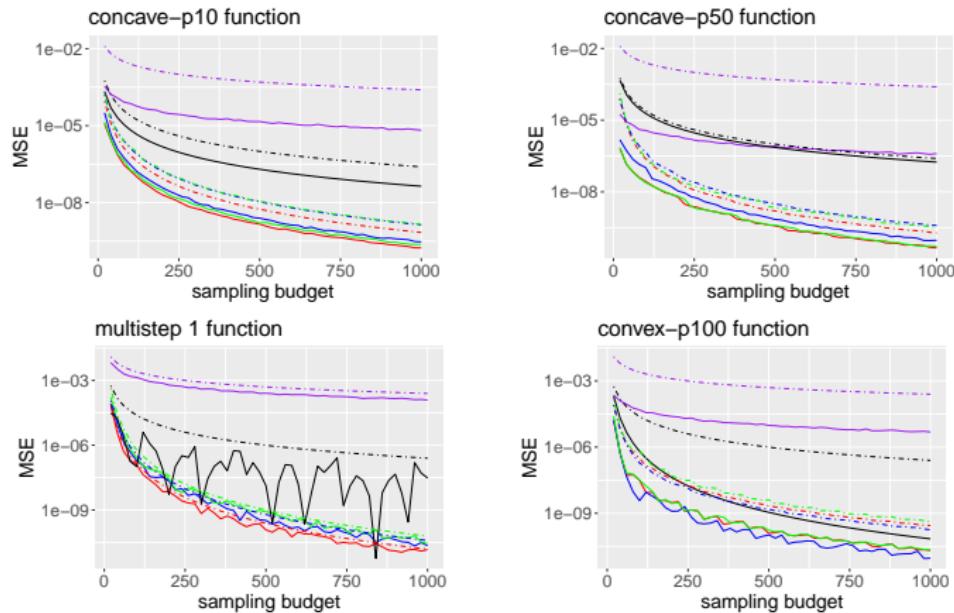


Figure 20: Empirical MSE (solid) and variance upper bound (dashed)
 → Methods : Trapezoidal, MC, Reference, [full_alloc](#), [full_alloc_init](#)

Summary conclusions

- Sequential methods are beneficial for **monotone** functions!
- **Compared to Monte Carlo** : proposed methods shows **significant benefit**
- **Compared to Trapezoidal rule**
 - Trapezoidal rule can be better in some cases (smooth func.)
 - Proposed methods outperform with increased **steepness** or **step** func.
- **Compared to Reference method (Novak, 1992)**
 - Proposed modifications are **always beneficial**

Key takeaways

- In class F of **monotone** functions (Novak, 1992)
 - **Sequential randomized** methods are beneficial!
 - Rate optimal w.r.t the order $n^{-3/2}$ of the L_1 error
- Benefits over **Trapezoidal rule** when function variance is concentrated
 - QMRA batch outputs: **step-like/stEEP** functions
- **Application QMRA:** Unknown distribution of integrating variable Y_x^{milk}
 - Distributional assumptions (e.g. *Listeria*)
 - Quantile estimates with reasonable simulation cost of Y_x^{milk} ?

C Appendix: Monotone Integration

Unknown distribution

- Transformation $x = F_Y(y)$ is used to reduce $\mathbb{E}(g(Y))$ into $\int_0^1 f(x)dx$

$$f = g \circ F_Y^{-1}$$

- When F_Y is unknown, $F_Y^{-1}(.)$ is replaced by empirical quantile estimates
- This induces an additional error in approximation of $\widehat{S}_n(f)$
- The algorithm remains same with the same optimal allocation rule
- Trade-off: cost of sampling of Y and evaluation of $g(.)$

Error bounds

- Nonseq. MC (Novak, 1992) Theorem 1: $e_1(\widehat{S}_n) \geq \frac{1}{8n}$
- Nonseq. MC (Basak et al., 2022b) Theorem 2.1 : $e_p(\widehat{S}_n) \geq \left(\frac{1}{2}\right)^{2+1/p} \frac{1}{n}$
- Sequential MC (Novak, 1992) Theorem 2: $e_1(\widehat{S}_n) \geq \frac{\sqrt{2}}{32} \cdot n^{-3/2}$
- Novak's sequential method (Novak, 1992) Theorem 3:

$$e_1(\widehat{S}_n^{\text{novak}}) \leq \sup_{f \in F} \text{Var}(\widehat{S}_n^{\text{novak}})^{1/2} \leq \frac{1}{\sqrt{2}} \cdot (m+1)^{-3/2} < \frac{\sqrt{54}}{2} \cdot n^{-3/2}$$
- Improved version of Novak's method:

$$e_1(\widehat{S}_n^{\text{novak.bis}}) < \frac{\sqrt{54}}{4} \cdot n^{-3/2}$$

Greedy sequential methods (an idea!)

- Greedy strategy can be beneficial for “step-like” functions
- Instead of **equispaced** initial strata split strata “greedily”
 - Using a dichotomy on strata bound difference Δ_k
 - Iteratively split the strata into half which has bigger Δ_k
- **Pro:** Allocates more budget to strata with high functional variance
- **Future work:** Establish worst case guarantees on error
- Preliminary numerical experiments show improvements on particular class of functions

Improved variance upper bound

For any monotone non-decreasing function $f \in F$, we have $\Delta_{k,i}^2 \geq 0$, which implies $\sum_{i=1}^{n_k} \Delta_{k,i}^2 \leq (\sum_i^{n_k} \Delta_{k,i})^2 = \Delta_k^2$. Thus we can write,

$$\begin{aligned}
 \text{Var}(\widehat{S}_n^{\text{novak}}(f)) &\leq \sum_{k=1}^{m+1} \frac{w_k^2}{n_k^2} \cdot \frac{1}{4} \sum_{i=1}^{n_k} \Delta_{k,i}^2, \quad \text{using inequality by Popoviciu (1935)} \\
 &\leq \frac{1}{4} \sum_{k=1}^{m+1} \frac{w_k^2 \Delta_k^2}{n_k^2} \\
 &\leq \frac{1}{4 \cdot (m+1)^3}, \quad w_k = \frac{1}{m+1} \& n_k \geq \Delta_k \cdot (m+1)
 \end{aligned} \tag{6}$$

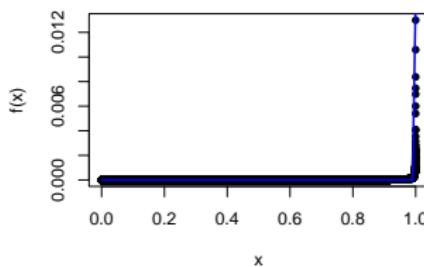
QMRA application bottle-necks

- Multivariate integration of $P^{\text{batch}}(\Xi)$
 - Only univariate integration problems are studied
- Partial monotonicity w.r.t Y_x^{milk}
 - Fix the stochastic variables $\Xi \setminus Y_x^{\text{milk}}$
 - OR
 - Integrate the output, for e.g. t^{consum}
- Unknown distribution Y_x^{milk}
 - Distributional assumption, like *Listeria*

QMRA Application

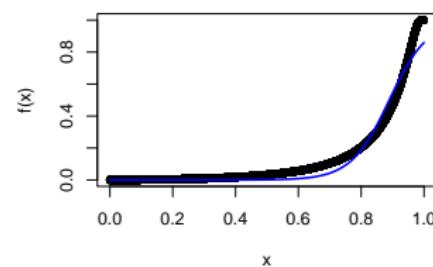
- The distribution of the integrating variable Y_x^{milk} is **unknown**
- Transformed $[0, 1] \rightarrow [0, 1]$ can be achieved using **distributional assumptions**

Fitting a Convex Function



(a) Approximated R_x^{batch} function

Fitting a Sigmoid Function



(b) Approximated P_x^{batch} function

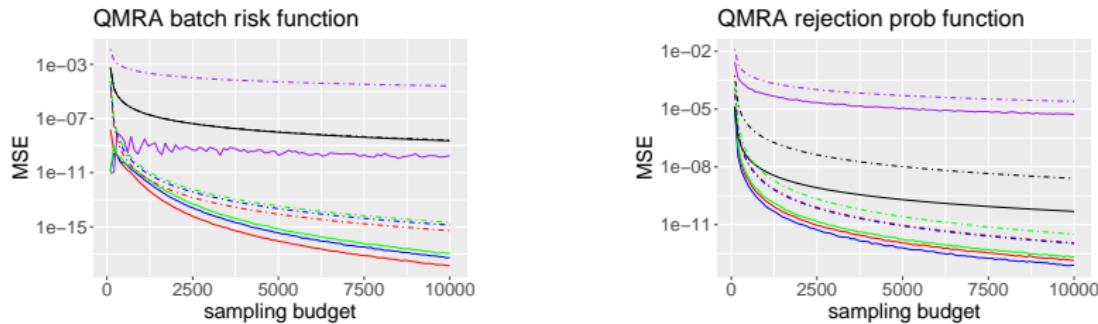


Figure 22: Empirical MSE (solid) and variance upper bound (dashed)

→ Methods : Trapezoidal, MC, Reference, full_alloc, full_alloc_init

- Sequential methods are beneficial in QMRA application problems
- Given assumptions on distribution of integrating variables

D Appendix: MOSO

Weights based on dominated area

Algorithm 2 Construction of $m(x)$ at step n

Estimate Pareto set $\widehat{\mathcal{P}}_n$ base on GP posterior mean

$\mathcal{F}_n^+ \leftarrow \{R_n^+(x) | x \in \widehat{\mathcal{P}}_n\}$ \triangleright (Pessimistic Pareto Front)

$V_{\text{ref}} \leftarrow D(\mathcal{F}_n^+)$ \triangleright (Reference dominated volume)

for $x \in \mathcal{C}$ **do**

$\mathcal{F}_n \leftarrow$ Pareto front of $\{\mathcal{F}_n^+ \cup R_n^-(x)\}$

$V_n \leftarrow D(\mathcal{F}_n^+)$ \triangleright (Dominated volume)

$m(x) \leftarrow (V_n - V_{\text{ref}}) / V_{\text{ref}}$ \triangleright (normalized in $[0, 1]$)

end for

Return $m(x)$

Weighted-MSE methods

- w-MSE- α

- $w_n(x) = \mathbb{1}_{\{x \in \Gamma_n^\alpha\}}$

$$\Gamma_n^\alpha = \{x | p_n(x) > \alpha \cdot \min_{x \in \widehat{\mathcal{P}}_n} p_n(x)\}$$

- Tested on $\alpha = 0.1, 0.5$

- w-MSE- λ

- $w_n(x) = \lambda \cdot p_n(x) + (1 - \lambda) \cdot p_n(x) \cdot (1 - p_n(x))$

- Tested on $\lambda = 0, 0.33, 1$

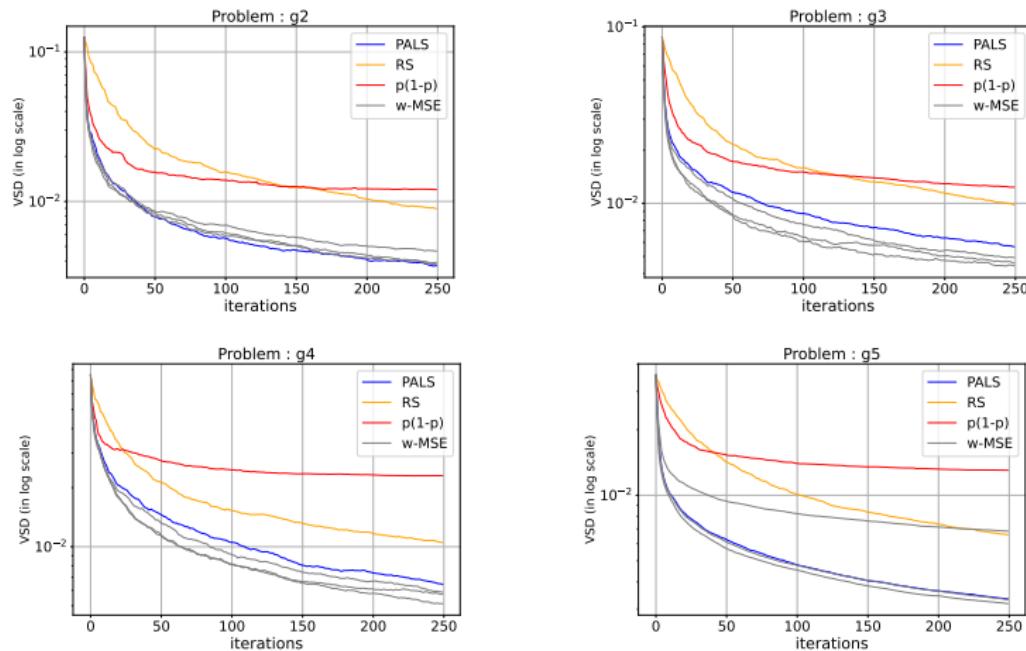


Figure 23: Average VSD metric on g-2,3,4,5 (Barracosa et al. (2021))

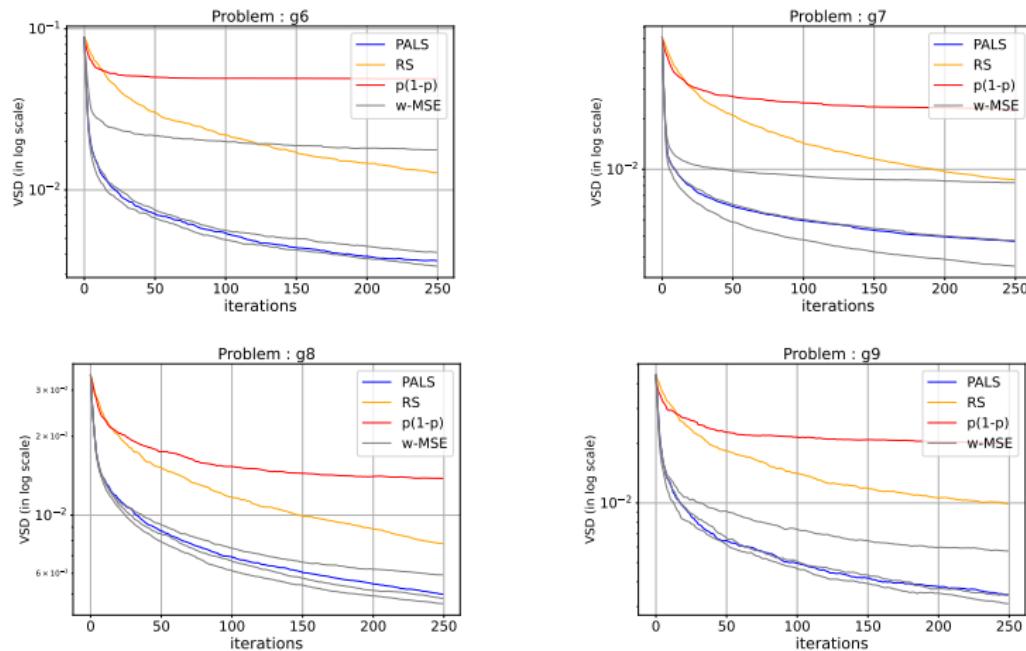


Figure 24: Average VSD metric on g-6,7,8,9 (Barracosa et al. (2021))

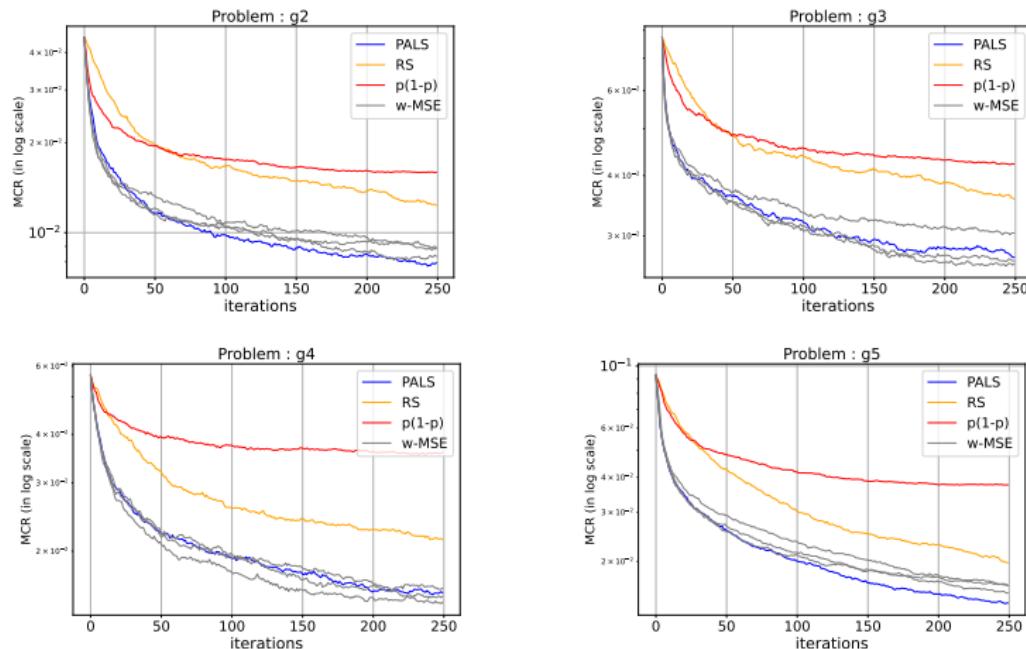


Figure 25: Average MCR metric on g-2,3,4,5 (Barracosa et al. (2021))

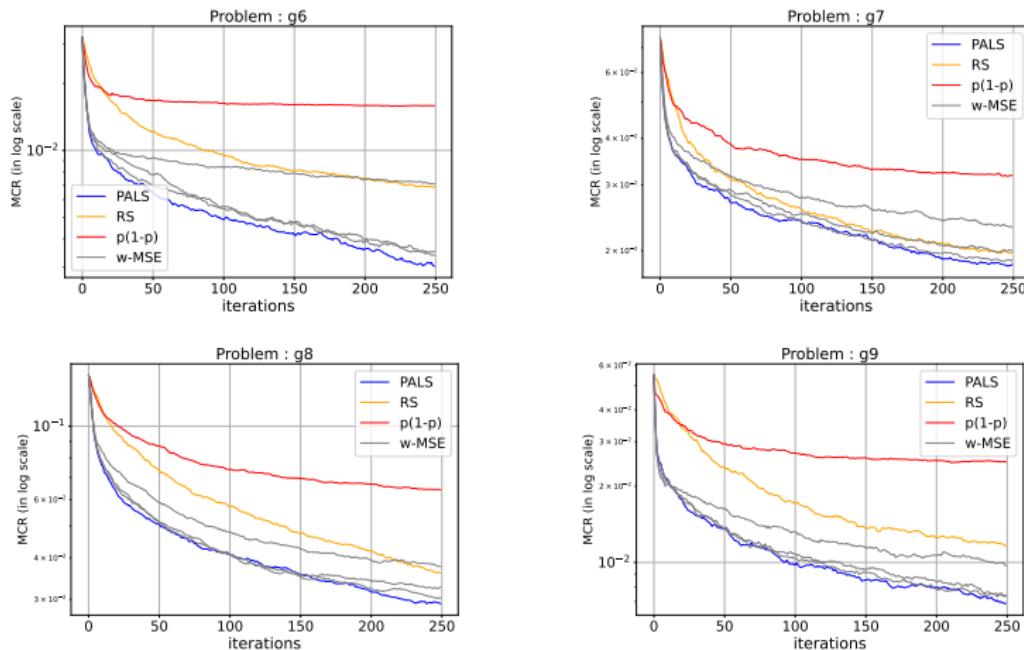


Figure 26: Average MCR metric on g-6,7,8,9 (Barracosa et al. (2021))

QMRA application - MPS-STEC model

- The Risk is not an expectation of the QMRA simulator outputs

$$R_{\text{MPS-STEC}} = \frac{\mathbb{E}[R^{\text{batch}}(1 - P^{\text{batch}}p^{\text{cheese}})]}{(1 - \mathbb{E}[P^{\text{batch}}p^{\text{cheese}}]) R_{\text{MPS-STEC}}^{\text{baseline}}}$$

- Required in a stochastic framework to perform batch evaluations
- Modified PALS algorithm (Basak et al., 2022a):
 - Fit GP surrogates on simulator outputs not objectives
 - Estimate quantiles of objectives using GP posteriors
 - Construct PALS rectangles using quantiles of the objectives

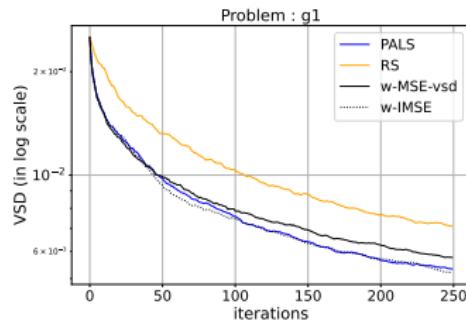
Acquisition function & choice of weights

- Choice of weights:
 - Sample the “Potentially Pareto optimal” points
 - $w_n^{\text{PALS}}(x_i) = \mathbb{1}_{\{x_i \in \mathbb{X} \setminus N_n\}}$
 - * Based on **symmetric** confidence rectangles
 - $w_n^{\text{Proposed}}(x_i) = \mathbb{1}_{\{x_i \in \mathbb{X} \setminus N_n\}} \cdot (V_{n,i} - V_n)$
 - * Extension of **PALS** weights w.r.t the **VSD** metric
- Acquisition functions:
 - **MUS**: Maximizing **local** uncertainty
 - (Pseudo) **SUR**: Minimizing **global** uncertainty

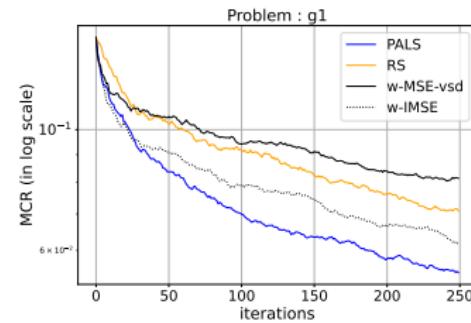
Key takeaways

1. $w_n^{\text{PALS}}(x_i) + \text{MUS} \rightarrow \text{Barracosa et al. (2021)}$
2. $w_n^{\text{PALS}}(x_i) + (\text{Pseudo}) \text{ SUR}$
 - **NO** significant improvement over 1.
3. $w_n^{\text{Proposed}}(x_i) + \text{MUS}$
 - Improvement on $\text{VSD}(\mathcal{F})$ metric over 1. & 4.
 - **Poor** performance on $\text{MCR}(\mathcal{P})$ metric
4. $w_n^{\text{Proposed}}(x_i) + (\text{Pseudo}) \text{ SUR}$
 - Improvement on $\text{VSD}(\mathcal{F})$ metric over 1.
 - **NO** improvement on $\text{MCR}(\mathcal{P})$ metric over 1.

Comparison between w-MSE-vsD, w-IMSE and PALS



(a) VSD metric



(b) MCR metric

- w-MSE-vsD improves slightly over w-IMSE on VSD metric
- w-MSE-vsD fails drastically on MCR metric

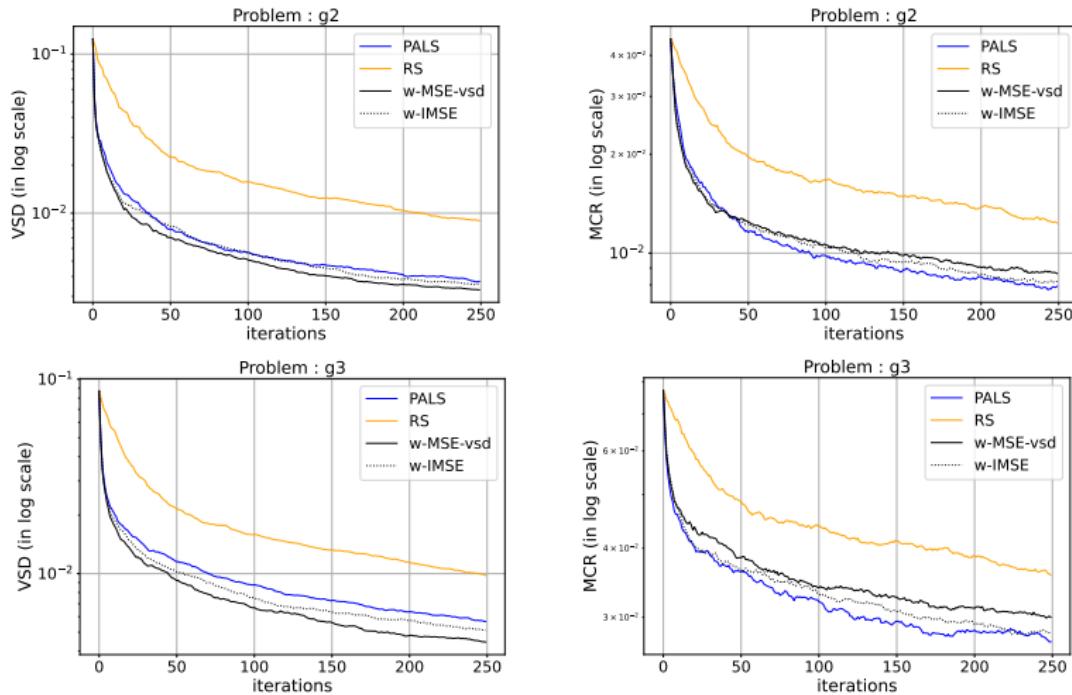


Figure 28: VSD metric (left) & MCR metric (right) for g2 and g3

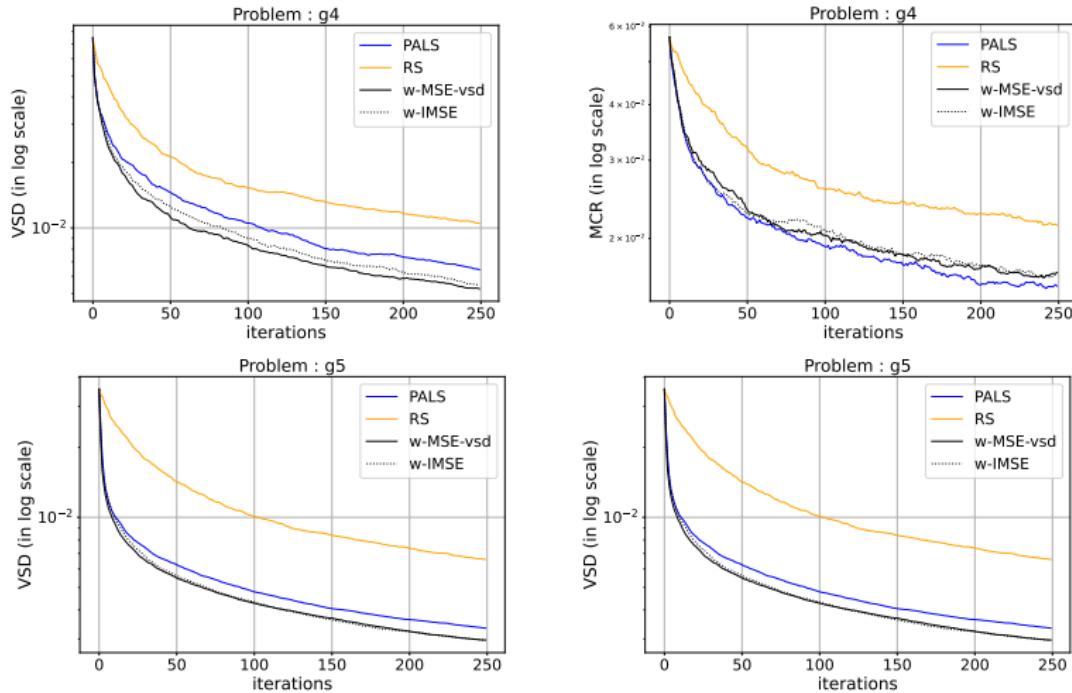


Figure 29: VSD metric (left) & MCR metric (right) for g4 and g5

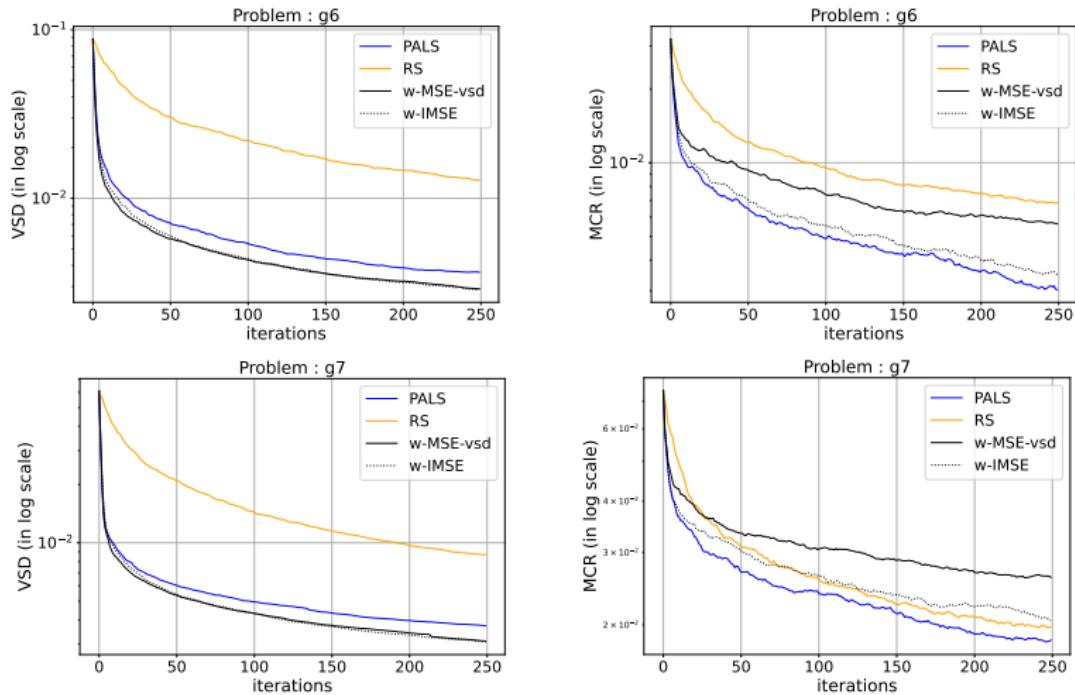


Figure 30: VSD metric (left) & MCR metric (right) for g6 and g7

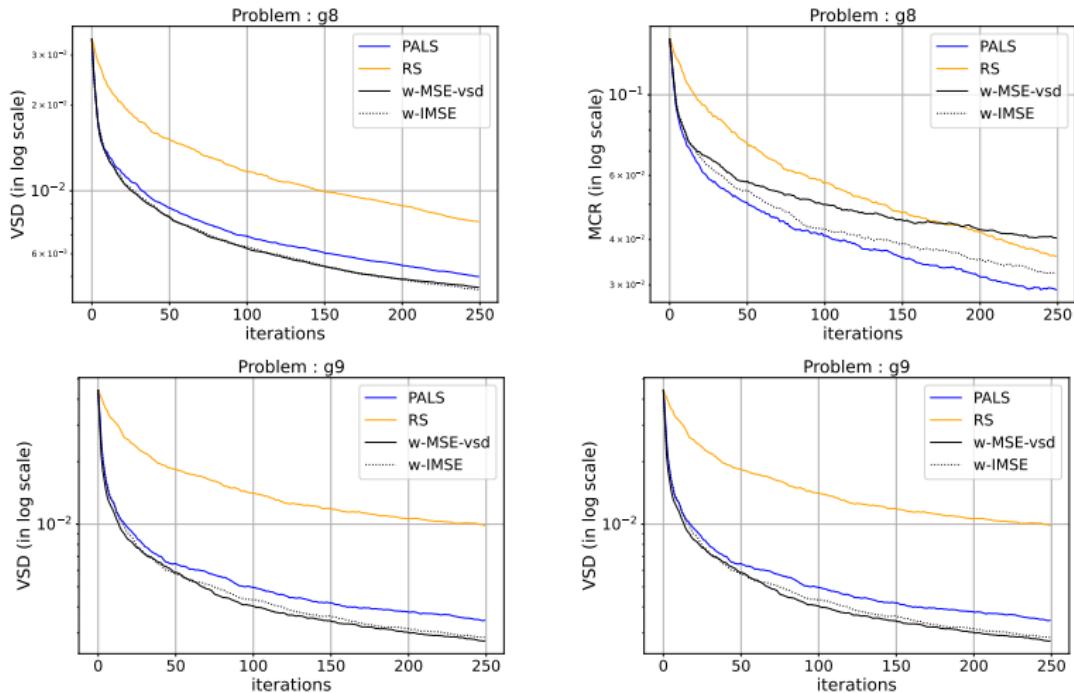


Figure 31: VSD metric (left) & MCR metric (right) for g8 and g9

References

- J. C. Augustin, V. Zuliani, M. Cornu, and L. Guillier. Growth rate and growth probability of listeria monocytogenes in dairy, meat and seafood products in suboptimal conditions. Journal of Applied Microbiology, 99:1019–1042, 2005. doi: 10.1111/j.1365-2672.2005.02710.x.
- 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.
- S. Basak, J. Bect, L. Guillier, F. Tenenhaus-Aziza, J. Christy, and E. Vazquez. Bayesian multi-objective optimization for quantitative risk assessment in microbiology. In MASCOT-NUM 2022, 2022a.
- 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, Jun 2022b.
- S. Basak, J. Bect, and E. Vazquez. Bayesian multi-objective optimization for stochastic simulators. In MASCOT-NUM 2023, 2023a.

- S. Basak, J. Christy, L. Guillier, F. Audiat-Perrin, M. Sanaa, F. Tenenhaus-Aziza, J. Bect, and E. Vazquez. Minimizing risk of illness and analytical costs using a qmra model for raw milk cheeses. In ICPMF 2023, 2023b.
- S. Basak, J. Christy, L. Guillier, F. Audiat-Perrin, M. Sanaa, F. Tenenhaus-Aziza, J. Bect, and E. Vazquez. Quantitative risk assessment of haemolytic and uremic syndrome (hus) from consumption of raw milk soft cheese. Food and Ecological Systems Modelling Journal, 5:e109502, 2024. doi: 10.3897/fmj.5.109502. URL <https://doi.org/10.3897/fmj.5.109502>.
- J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. Statistics and Computing, 2011.
- S. Belakaria, A. Deshwal, and J. R. Doppa. Max-value entropy search for multi-objective bayesian optimization with constraints, 2020.
- M. Binois, D. Ginsbourger, and O. Roustant. Quantifying uncertainty on pareto fronts with gaussian process conditional simulations. European Journal of Operational Research, 2015.
- L. Bonifait, A. Thépault, L. Baugé, S. Rouxel, F. Le Gall, and M. Chemaly. Occurrence of salmonella in the cattle production in france. Microorganisms, 9(4):872, 2021.
- B. Bryan, R. C. Nichol, C. R. Genovese, J. Schneider, C. J. Miller, and L. Wasserman. Active learning for identifying function threshold boundaries. In Advances in Neural Information Processing Systems, 2005.

A. Cassini, E. Colzani, A. Pini, M. J. J. Mangen, D. Plass, S. A. McDonald, G. Maringhini, A. van Lier, J. A. Haagsma, A. H. Havelaar, et al. Impact of infectious diseases on population health using incidence-based disability-adjusted life years (dalys): results from the burden of communicable diseases in europe study, european union and european economic area countries, 2009 to 2013. *Eurosurveillance*, 23(16):17-00454, 2018.

Codex Alimentarius Commission. Principles and guidelines for the conduct of microbiological risk assessment. CAC/GL-30, 1999.

- A. Fares. Quantitative risk assessment model of human salmonellosis linked to the consumption of Camembert cheese made from raw milk. PhD thesis, AgroParisTech, 2007.
- U. Gonzales-Barron, V. Cadavez, A. Valero, P. Skandamis, S. Kintzios, A. de Cesare, G. Manfreda, F. Tenenhaus-Aziza, L. Guillier, N. Boudhrioua, and F. Achemchem. Report on the First Predictive Dynamic Models of the Viability of Pathogens along Processing of Mediterranean Artisanal Fermented Foods and Report on the Optimised Process Variables to Enhance their Microbiological Safety, aug 2022. URL <https://doi.org/10.5281/zenodo.8118475>.
- J. M. Hernández-Lobato, M. W. Hoffman, and Z. Ghahramani. Predictive entropy search for efficient global optimization of black-box functions. In Ghahramani Z., Welling M., Cortes C., Lawrence N., and Weinberger K. Q., editors, Advances in Neural Information Processing Systems, volume 27. Curran Associates, Inc., 2014. URL <https://proceedings.neurips.cc/paper/2014/file/069d3bb002acd8d7dd095917f9efe4cb-Paper.pdf>.

- J. Kiefer. Optimum sequential search and approximation methods under minimum regularity assumptions. *Journal of the Society for Industrial and Applied Mathematics*, 5(3):105–136, 1957.
- L. H. Lee, E. P. Chew, S. Teng, and D. Goldsman. Finding the non-dominated pareto set for multi-objective simulation models. *IIE Transactions*, 42(9):656–674, 2010. doi: 10.1080/07408171003705367.
- E. Novak. Quadrature formulas for monotone functions. *Proceedings of the American Mathematical Society*, 115(1):59–68, 1992.
- F. Perrin, F. Tenenhaus-Aziza, V. Michel, S. Miszczycza, N. Bel, and M. Sanaa. Quantitative risk assessment of haemolytic and uremic syndrome linked to o157:h7 and non-o157:h7 shiga-toxin ProducingEscherichia coliStrains in raw milk soft cheeses. *Risk Analysis*, 35(1):109–128, 2014.
- T. Popoviciu. Sur l'approximation des fonctions convexes d'ordre superieur. *Mathematica (Cluj)*, 10:49–54, 1935.
- R. Pouillot, K. Hoelzer, Y. Chen, and S. B. Dennis. Listeria monocytogenes dose response revisited—incorporating adjustments for variability in strain virulence and host susceptibility. *Risk Analysis*, 35(1):90–108, 2015.
- R. Pouillot, A. Kiermeier, L. Guillier, V. Cadavez, and M. Sanaa. Updated parameters for listeria monocytogenes dose-response model considering pathogen virulence and age and sex of consumer. *Foods*, 13(5):751, 2024.

- A. Ricci, A. Allende, D. Bolton, M. Chemaly, R. Davies, P. S. Fernandez Escamez, R. Girones, L. Herman, K. Koutsoumanis, B. Norrung, L. Robertson, G. Ru, M. Sanaa, M. Simmons, P. Skandamis, E. Snary, N. Speybroeck, B. Ter Kuile, J. Threlfall, H. Wahlstrom, J. Takkinen, M. Wagner, D. Arcella, M. T. Da Silva Felicio, M. Georgiadis, W. Messens, R. Lindqvist, and EFSA Panel on Biological Hazards (BIOHAZ). Listeria monocytogenes contamination of ready-to-eat foods and the risk for human health in the eu. *EFSA Journal*, 16(1), 2018. doi: <https://doi.org/10.2903/j.efsa.2018.5134>. URL <https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2018.5134>.
- S. Rojas Gonzalez, H. Jalali, and I. Van Nieuwenhuyse. A multiobjective stochastic simulation optimization algorithm. *European Journal of Operational Research*, 284(1):212–226, 2020. doi: 10.1016/j.ejor.2019.12.01.
- M. Sanaa, L. Coroller, and O. Cerf. Risk assessment of listeriosis linked to the consumption of two soft cheeses made from raw milk: Camembert of normandy and brie of meaux. *Risk Analysis: An International Journal*, 24(2):389–399, 2004.
- A. J. Strickland, F. Sampedro, and C. W. Hedberg. Quantitative risk assessment of salmonella in ground beef products and the resulting impact of risk mitigation strategies on public health. *Journal of food protection*, 86(6), 2023.
- P. F. M. Teunis, F. Kasuga, A. Fazil, I. D. Ogden, O. Rotariu, and N. J. C. Strachan. Dose–response modeling of salmonella using outbreak data. *International journal of food microbiology*, 144(2):243–249, 2010.

-
- E. Vazquez and J. Bect. A sequential bayesian algorithm to estimate a probability of failure. IFAC Proceedings Volumes, 42(10):546–550, 2009.
 - E. Vazquez and M. P. Martinez. Estimation of the volume of an excursion set of a gaussian process using intrinsic kriging. arXiv preprint math/0611273, 2006.
 - J. Villemonteix, E. Vazquez, and E. Walter. An informational approach to the global optimization of expensive-to-evaluate functions, 2007.
 - WHO. Microbiological Risk Assessment–Guidance for food, volume 36. Food & Agriculture Org., 2021.
 - M. Zuluaga, G. Sergent, A. Krause, and M. Puschel. Active learning for multi-objective optimization. In Proceedings of the 30th International Conference on Machine Learning, pages 462–470. PMLR, 2013.