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

Subhasish Basak

université PARIS-SACLAY

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

Appréciation quantitative des risques multipathogènes dans le fromage à pâte molle au lait cru, intégration monotone et optimisation Bayésienne.

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°580 Sciences et Technologies de l'Information et de la Communication (STIC)
Spécialité de doctorat: Informatique mathématique
Graduate School : Informatique et sciences du numérique, Référent : CentraleSupélec

Thèse préparée dans l'unité de recherche **Laboratoire des signaux et systèmes** (Université Paris-Saclay, CNRS, CentraleSupélec), sous la direction de **Emmanuel VAZQUEZ**, Professeur des Universités, le co-encadrement de **Julien BECT**, Maître de Conférence, **Laurent GUILLIER**, Chef de projets scientifiques et **Fanny TENENHAUS-AZIZA**, Directrice Data et Statistiques.

Thèse soutenue à Paris-Saclay, le 20 mars 2024, par

Subhasish BASAK

Composition du jury

Membres du jury avec voix délibérative

Pierre BARBILLON Professeur, AgroParisTech - Université Paris-Saclay	Président
Olivier ROUSTANT Professeur des universités, INSA Toulouse	Rapporteur
Jan VAN IMPE Professeur, Katholieke Universiteit Leuven	Rapporteur
Anne THEBAULT Chargée de recherche, ANSES	Examinatrice
Jeanne-Marie MEMBRÉ Ingénieure agronome, INRAE/ONIRIS Nantes	Examinatrice

THÈSE DE DOCTORAT

NNT : 2024UPASG021

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre: Appréciation quantitative des risques multipathogènes dans le fromage à pâte molle au lait cru, intégration monotone et optimisation Bayésienne.

Mots clés: Appréciation quantitative des risques, stochastique simulateur, intégration bornée monotone, optimisation Bayésienne.

Résumé: Ce manuscrit se concentre sur l'optimisation Bayésienne d'un modèle d'appréciation quantitative des risques microbiologiques (AQRM) dans le cadre du projet ArtiSaneFood soutenu par l'Union européenne. L'objectif est d'établir des stratégies de bio-intervention efficaces pour les fabricants de fromage au lait cru en France, en s'appuyant sur trois types de travaux : 1) le développement d'un modèle AQRM multipathogène pour un fromage de type pâte molle au lait cru, 2) étudier des méthodes d'intégration monotone pour l'estimation des sorties du modèle AQRM et 3) la conception d'un algorithme d'optimisation Bayésienne adapté à un simulateur stochastique et coûteux.

Dans la première partie, nous proposons un modèle AQRM multipathogène construit sur la base d'études existantes (voir, par exemple, Bonifait et al., 2021, Perrin et al., 2014, Sanaa et al., 2004, Strickland et al., 2023). Ce modèle est conçu pour estimer l'impact des maladies d'origine alimentaire sur la santé publique, causées par des agents pathogènes tels que *Escherichia coli* entérohémorragiques (EHEC), *Salmonella* et *Listeria monocytogenes*, potentiellement présents dans le fromage de type pâte molle au lait cru. Ce modèle "farm-to-fork" intègre les mesure de maitrise liées aux tests microbiologiques du lait et du fromage, permettant d'estimer les coûts associés aux interventions. Une implémentation du modèle AQRM pour EHEC est fournie en R et dans le cadre FSKX (Basak et al., 2024).

La deuxième partie de ce manuscrit explore l'application potentielle de méthodes d'intégration séquentielle, exploitant les propriétés de monotonie et de bornage des sorties du simulateur. Nous menons une revue de littérature approfondie sur les méthodes d'intégration existantes (voir, par exemple, Kiefer, 1957, Novak, 1992), et examinons les résultats théoriques concernant leur convergence. Notre contribution comprend la proposition d'améliorations à ces méthodes et la discussion des défis associés à leur application dans le domaine de l'AQRM.

Dans la dernière partie de ce manuscrit, nous proposons un algorithme Bayésien d'optimisation multiobjectif pour estimer les entrées optimales de Pareto d'un simulateur stochastique et coûteux en calcul. L'approche proposée est motivée par le principe de "Stepwise Uncertainty Reduction" (SUR) (voir, par exemple, Vazquez and Bect, 2009, Vazquez and Martinez, 2006, Villemonteix et al., 2007), avec un critère d'échantillonnage basé sur weighted integrated mean squared error (w-IMSE). Nous présentons une évaluation numérique comparant l'algorithme proposé avec PALS (Pareto Active Learning for Stochastic simulators) (Barracosa et al., 2021), sur un ensemble de problèmes de test bi-objectifs. Nous proposons également une extension (Basak et al., 2022a) de l'algorithme PALS, adaptée au cas d'application de l'AQRM.

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

Keywords: Quantitative risk assessment, stochastic simulator, Monotone bounded integration, Bayesian optimization.

Abstract: This manuscript focuses on Bayesian optimization of a quantitative microbiological risk assessment (QMRA) model, in the context of the European project ArtiSaneFood, supported by the PRIMA program. The primary goal is to establish efficient bio-intervention strategies for cheese producers in France. This work is divided into three broad directions: 1) development and implementation of a multipathogen QMRA model for raw milk soft cheese, 2) studying monotone integration methods for estimating outputs of the QMRA model, and 3) designing a Bayesian optimization algorithm tailored for a stochastic and computationally expensive simulator.

In the first part we propose a multipathogen QMRA model, built upon existing studies in the literature (see, e.g., Bonifait et al., 2021, Perrin et al., 2014, Sanaa et al., 2004, Strickland et al., 2023). This model estimates the impact of foodborne illnesses on public health, caused by pathogenic STEC, *Salmonella* and *Listeria monocytogenes*, which can potentially be present in raw milk soft cheese. This farm-to-fork model also implements the intervention strategies related to mlik and cheese testing, which allows to estimate the cost of intervention. An implementation of the QMRA model for STEC is provided in R and in the FSKX framework (Basak et al., 2024).

The second part of this manuscript investigates the potential application of sequential integration methods, leveraging the monotonicity and boundedness properties of the simulator outputs. We conduct a comprehensive literature review on existing integration methods (see, e.g., Kiefer, 1957, Novak, 1992), and delve into the theoretical findings regarding their convergence. Our contribution includes proposing enhancements to these methods and discussion on the challenges associated with their application in the QMRA domain.

In the final part of this manuscript, we propose a Bayesian multiobjective optimization algorithm for estimating the Pareto optimal inputs of a stochastic and computationally expensive simulator. The proposed approach is motivated by the principle of Stepwise Uncertainty Reduction (SUR) (see, e.g., Vazquez and Bect, 2009, Vazquez and Martinez, 2006, Villemonteix et al., 2007), with a weighted integrated mean squared error (w-IMSE) based sampling criterion, focused on the estimation of the Pareto front. A numerical benchmark is presented, comparing the proposed algorithm with PALS (Pareto Active Learning for Stochastic simulators) (Barracosa et al., 2021), over a set of bi-objective test problems. We also propose an extension (Basak et al., 2022a) of the PALS algorithm, tailored to the QMRA application case.

Remerciements

Je tiens à remercier sincèrement tous ceux qui ont contribué, d'une manière ou d'une autre, à l'organisation, au développement et à la réalisation de cette thèse.

Tout d'abord, je tiens à remercier le directeur de ma thèse, Emmanuel Vazquez, de m'avoir accueilli en tant que stagiaire en 2019 et ensuite d'avoir accepté ma candidature en tant que doctorant, de m'avoir aidé à m'installer en France au milieu de la pandémie et surtout, pour sa disponibilité, ses précieux conseils, ses encouragements, sa bienveillance et sa gentillesse tout au long de ma thèse.

Je dois énormément de gratitude à mon encadrant, Julien Bect. Travailler avec lui a été une expérience enrichissante de ma vie. Je le remercie pour sa disponibilité remarquable, ses conseils, les échanges d'idées, les sessions de discussion au tableau, et surtout pour m'avoir enseigné l'organisation au travail, la précision et la rigueur scientifique.

Je tiens à remercier mon encadrant, Laurent Guillier de l'ANSES, pour son soutien, sa bienveillance, son encadrement, pour m'avoir introduit au domaine de la sécurité des aliments, pour m'avoir aidé à réseauter avec des experts du monde entier, et pour m'avoir guidé dans le choix de ma prochaine étape dans ma carrière professionnelle.

Un grand merci à Fanny Tenenhaus-Aziza, car cette thèse n'aurait pas été réalisable sans sa contribution, sa coordination et son soutien. Je tiens à exprimer ma gratitude pour m'avoir accueilli à la Maison du Lait à Paris et de m'avoir fait découvrir l'univers des fromages français. Je tiens également à exprimer ma gratitude pour son encadrement, sa bienveillance, son soutien moral et nos conversations.

Je souhaite adresser mes sincères remerciements à Pierre Barbillon pour avoir présidé le jury de la soutenance de ma thèse, à Olivier Roustant et Jan Van Impe pour m'avoir fait l'honneur d'être rapporteurs de cette thèse, ainsi qu'à Anne Thebault et Jeanne Marie Membré pour avoir accepté d'en être examinatrices. Je tiens également à remercier les membres du jury de mi-soutenance de thèse, Jean-Christophe Augustin et Jean-Philippe Ovarlez, pour leur expertise et leurs conseils.

Ensuite, je tiens à remercier tous les partenaires et collègues du projet ArtiSaneFood, en France et dans les autres pays participants, avec lesquels j'ai eu l'occasion de rencontrer, d'échanger des idées et de travailler. Un merci particulier à Janushan Christy d'ACTALIA pour sa contribution, ainsi qu'à Éloïse Modric du Fromage AOP de Normandie pour nous avoir accueillis en Normandie. Je remercie également les collègues de l'ANSES et du CNIEL avec lesquels j'ai eu l'occasion d'échanger pendant ma période de thèse : Moez Sanaa, Frédérique Audiat-Perrin, Anne, Du, Juliana et chère amie Dihia.

Pendant ces trois années de thèse, le L2S m'a offert un environnement de travail particulièrement agréable, au sein duquel j'ai eu le plaisir de rencontrer des collègues formidables. Je commence par remercier les permanents du labo : Arthur, Laurent, Gilles, José, Elisabeth et Charles. Ensuite, les doctorants qui étaient avec moi depuis le début, Sébastien et Fabien, qui m'ont toujours inspiré, ont enrichi mes connaissances grâce à des discussions fructueuses et étaient toujours présents pour m'aider sur n'importe quel sujet. Merci également à Lucas, Jean et Elen qui ont toujours su égayer l'ambiance avec leur bonne humeur, et qui m'ont appris plein de trucs intéressants, comme l'histoire de France, la politique, la gastronomie française, le code parisien et surtout les gros mots français, haha! Je remercie aussi tous les autres doctorants, post-doc et stagiaire que j'ai eu le plaisir de ren-

contrer et de partager des moments avec : Romain, Manon, Bruno, Eniko, Reihan, Aurélien, Sarah, Nathan, Théo et Adama. Et enfin, un grand merci à Lisa pour ses visites de temps en temps et pour tous ces moments sympas qu'on a passé ensemble.

Merci également à mes collègues de l'ANSES de Lyon pour leur bienveillance, nos échanges, leur soutien et leur assistance sur les démarches administratives tout au long des trois derniers mois de ma thèse hors Paris. Un grand merci à Jean-Philippe, Viviane, Lucie, Nunzio, l'équipe EAS et à mes collègues avec qui j'ai partagé mon bureau : Morgane, Leila, Servane, Maissane et cara collega Antonella.

Je tiens encore une fois à remercier chaleureusement Fanny, ainsi qu'Aurthur et leurs enfants pour leur incroyable gentillesse, leurs délicieux repas et les super moments passés ensemble.

Il est impossible que j'oublie ma famille et mes amis en Inde. Un grand merci à tous !

À mes parents, pour leur amour inconditionnel, leur soutien et surtout pour la liberté qu'ils m'ont toujours accordée dans mes choix.

Je dédie ce travail à mon grand-père qui a toujours été une source d'inspiration non seulement pour moi mais aussi pour tout les membres de ma famille. Je réalise toute la chance que j'ai eu de vivre une vie si prospère grâce à son sacrifices, son talent et son travail acharné.

“If you are lucky enough to have lived in Paris as a young man, then wherever you go for the rest of your life, it stays with you, for Paris is a moveable feast.”
— Hemingway (1964)

Contents

1 - Introduction

According to the World Health Organization (WHO), around the world, an estimated 600 million—almost 1 in 10 people—fall ill after consuming contaminated food each year, resulting in $420,000$ deaths and a loss of 33 million disability-adjusted life years (DALYs). DALYs, a societal measure of the disease or disability burden in populations, representing the combined years lost to premature death and years lived with disability, serve as a vital indicator for quantifying the profound health consequences of food-borne illnesses (World Health Organization et al., 2015). Understanding the significant impact of DALYs, it becomes evident that food safety is not only essential for human health and well-being but also for promoting sustainable development, economic advancement, and enhanced productivity. The 2019 World Bank report on the economic burden of food-borne diseases indicated that USD 110 billion is lost each year in productivity and medical expenses due to unsafe food in low- and middle-income countries. Typically, food-borne illnesses are either infectious or toxic in nature and are caused by the intake of bacteria, viruses, parasites, or chemical substances into the body through contaminated food. In this context, microbiological food safety emerges as a scientific discipline to study the set of measures and practices aimed at preventing food-borne illnesses. Ensuring food safety and implementing efficient prevention measures have been major challenges for the food sector. Great research efforts have been devoted to addressing these challenges by the World Health Organization (WHO), the European Food Safety Agency (EFSA), various governmental agencies.

1.1 . Context

This thesis is part of the European project ArtiSaneFood, which focuses on innovative bio-interventions and risk modeling approaches to ensure microbial safety and quality of Mediterranean artisanal fermented foods. ArtiSaneFood has a budget of EUR 1.58 million and started in June 2019. It involves cooperation between Mediterranean countries, namely Portugal, Spain, France, Italy, Greece, Morocco and Tunisia. For France, an objective is to optimize intervention strategies, such as milk sorting or product testing, during the production of raw milk cheeses. Recognized by the UNESCO as a unique cultural heritage, raw milk cheeses, as shown in Figure 1.1, remain an integral part of the French gastronomy.

However, the consumption of raw milk cheeses can pose a risk of food-borne illnesses to the consumers, due to the presence of pathogenic bacterias

Figure 1.1: Un plateau de fromage français. Photo credit: ANSES.

in it. According to an opinion (Dubois-Brissonnet et al., 2022) published by the French food safety agency ANSES, over the last decade in France, 34%, 37%, and 60% of outbreaks of salmonellosis, listeriosis, and enterohaemorrhagic *E. coli* (EHEC) infections respectively have been linked to the consumption of raw-milk cheeses. While some bacteria, such as *Salmonella spp.* and *Staphylococcus aureus*, can cause gastroenteritis symptoms, others can have much more serious consequences, such as kidney failure (EHEC) or even death (*L. monocytogenes*, EHEC). The main sources of these hazards are soft cheeses with a surface mould and short-ripened uncooked pressed cheeses.

To control the risk linked to the consumption of raw milk cheese, several control measures are implemented in practice at different stages of the cheese production chain. Following Perrin et al. (2014), these intervention measures can be categorized into two types: namely, preharvest and postharvest interventions. The former pertains to the presence of pathogens in the farm milk used for cheese production by monitoring/improving the hygiene conditions of milk-producing farms. Some of the measures practiced at the farm level include the detection and isolation of high shedders of pathogens, vaccination, use of probiotics, antimicrobials, bacteriophages, or altering diet (see, e.g., Farrokh et al., 2013). At the production level, milk acquired from farms undergoes microbiological testing to detect potential contamination of pathogens, resulting in the removal of milk surpassing a specific contamination threshold; this practice is known as milk sorting. Postharvest interventions are introduced all along the production process, focusing on microbiological testing of the products and the environment of the production. This includes sampling test units from particular batches of cheese during or after production and testing them for potential pathogen contamination. Generally if one sample unit tests positive for pathogen contamination (e.g. MPS-STEC, *Listeria monocytogenes* or *Salmonella*), the batch of cheese is rejected, preventing its entry into the market for sale. Clearly, intervention strategies are necessary for controlling the microbiological risk linked to the consumption of raw milk cheese, but strict intervention strategies can also lead to significant loss of raw milk and final cheese products, imposing an economic burden on the

cheese producers. In other words, the reduction of the risk of food-borne illness is also linked to the cost of the implementation of strict control measures of hygiene and the rejection of contaminated food items from the production process. This trade-off presents a problem of implementing an efficient and economically acceptable intervention strategy for the cheese producers, in light of microbiological food safety. More precisely, the objective is to find optimal process intervention parameters which, "in some sense", are able to minimize simultaneously the risk of food-borne illnesses and the cost of intervention strategies.

To address this multidisciplinary problem, the ArtiSaneFood project brings together several French organizations. This consortium includes Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (ANSES), Centre National Interprofessionnel de l'Economie Laitiere (CNIEL), Centre technique d'expertise agroalimentaire (ACTALIA), Laboratoire des Signaux et Systèmes (L2S), CNRS, CentraleSupélec, Université Paris-Saclay, Conseil national des appellations d'origine laitières (CNAOL), and Fromages AOP de Normandie. Collectively, these organizations provide specific expertises in distinct domains, including microbiological risk assessment, the design and analysis of computer experiments, and research and development within the French dairy industry, to oversee and guide this project.

1.2 . Academic Context and Directions of Research

In this thesis, we formalize the aforementioned objective from a mathematical point of view using statistical and optimization tools. Before any analysis or optimization, it is necessary to model the evolution of the pathogens in raw milk soft cheese to estimate suitable metrics to assess the risk of illness and the cost of intervention. In the field of microbiology for the agri-food sector, Quantitative Microbiological Risk Assessment (QMRA) is a scientific discipline devoted to analyzing microbiological risks, examining the links between producer and consumer practices and the risks of food contamination (World Health Organization et al., 2021). The microbiological quality of a food product results from numerous operations throughout the milk processing, cheese manufacturing and preservation process, controlled by several parameters including temperatures, duration, flow rates, pH, and intervention measures, among others. QMRA models utilize differential equations with random initial conditions to model the growth of pathogens under dynamic environmental conditions, rendering these models stochastic as they are contingent upon random quantities and several dependent stochastic internal variables. Statistical analyses derived from these stochastic models and simulations are then used to measure the impact of intervention measures on the risk of illness and cost.

In the realm of QMRA for raw milk soft cheese, the state-of-the-art model was provided by Perrin et al. (2014), evaluating the effect of pre- and postharvest interventions on public health risks associated solely with Shiga toxin-producing *Escherichia coli*. However this single-pathogen QMRA model did not consider the effects of other pathogens, namely, *Listeria monocytogenes* and *Salmonella*, which can be potentially present in raw milk cheese (see, e.g., Costanzo et al., 2020, Sanaa et al., 2004, Tenenhaus-Aziza et al., 2014). Moreover, Perrin et al. (2014) employed a scenario-based statistical analysis to assess the impacts of various intervention methods, a technique commonly adopted in QMRA literature. But this "classical" approach becomes impractical when simulation costs are high, and it is only suited for comparing different intervention scenarios in terms of risk reduction.

In some instances, Bayesian decision-theoretic methods have been employed by certain studies to optimize specific process control measures (see, e.g., Commeau, 2012), but to our knowledge, the existing body of QMRA literature lacks instances where optimal process intervention parameters are identified to simultaneously minimize two or more conflicting outcomes.

We place ourselves within the framework of Multiobjective Simulation Optimization (MOSO) for functions that are expensive to evaluate (see, e.g., Barracosa et al., 2021, Frazier et al., 2009, Hunter et al., 2019). To address this problem, we use the principle of Bayesian optimization (Jones et al., 1998). Consider a set $\mathcal{F}$ of real-valued functions defined on a set $\mathbb{X} \subseteq \mathbb{R}^p$ mapping onto the set $\mathbb{R}^q$, and let $\phi : \mathcal{F} \to \mathcal{G}$ be a given mapping. The objective is to optimize the quantities of interest $\phi(f)$, using a finite set of (potentially noisy) evaluations of f, where the real-valued function $f \in \mathcal{F}$ is computationally expensive to evaluate, and the analytical form of f is either unknown or too complex to use a classical derivative-based approach for optimization. Within the framework of Bayesian decision theory, the evaluations of f are chosen adaptively; the experimenter sequentially observes and decides at each stage which new evaluation should be performed, based on the information collected up to that stage. To infer about $\phi(f)$, at the n-th stage, an estimator $\hat{\phi}_n \in \mathcal{G}$ is constructed, typically using a Gaussian process (Rasmussen and Williams, 2006) surrogate model, based on the observations up to the n-stage. Additionally, a sampling criterion $J_n : \mathbb{X} \to \mathbb{R}$ is computed, based on $\hat{\phi}_n$, to decide the next evaluation point $X_{n+1} \in \mathbb{X}$.

In our framework, the QMRA simulator can be formalized as a function $f \in \mathcal{F} : \mathbb{X} \to \mathbb{R}^q$ with q noisy outputs and the simulator input space $\mathbb{X}$, consisting of the process intervention parameters. Then, the problem of multiobjective optimization becomes estimating the solution set of $\operatorname{argmin}_{\mathbb{X}} f_1, f_2, \ldots, f_q$, using finite evaluations of f. The trade-off between conflicting objectives is reconciled by the concept of Pareto optimality. A set of solutions $\mathcal{P} \subset \mathbb{X}$ is deemed Pareto optimal if there exists no alternative solution $x' \in \mathbb{X}$ that offers im-

provements for all objectives without compromising the performance of at least one objective.

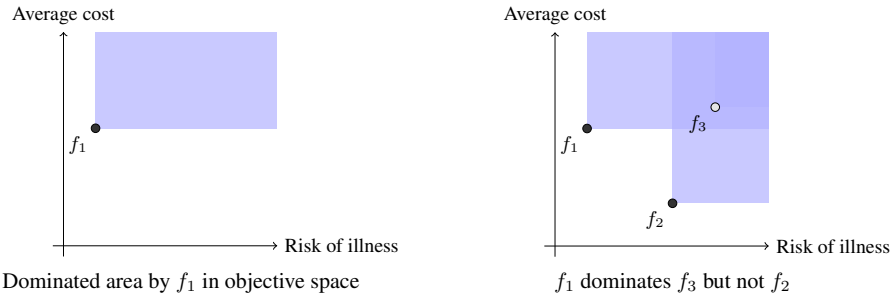


Figure 1.2: Notion of Pareto optimality demonstrated with three evaluations and corresponding dominated regions.

In Figure 1.2, the shaded regions represent the area dominated by that particular point, meaning any other point falling in that region will have either a higher risk or higher cost and will thus be considered dominated. Following this principle, the decision-maker would choose the points f_1 or f_2 over f_3 to minimize the outputs with respect to both objectives. In this case, both f_1 and f_2 constitute the Pareto optimal set $\mathcal{P}$, as they are optimal with respect to one of the objectives, and it is not possible to minimize one objective without deteriorating the other. The problem becomes more challenging in the presence of noise, particularly when the f_is are observed with noise due to the stochastic nature of the simulator. For our problem, we aim to find the set of optimal process intervention parameters for the QMRA simulator, such that the resulting set of risk and cost will represent the best possible choice of trade-off.

Concerning the choice of parameters to be optimized, although the literature does present some works employing sensitivity analysis on different process parameters to identify those having the greatest impact (see, e.g., Duret et al., 2014, Lamboni et al., 2014), the primary focus of this thesis is to offer methodological recommendations exclusively for the pre-identified intervention parameters. More precisely, we are interested in the process intervention parameters that control the impact of the intervention, for example, the frequency of testing the farm milk, the threshold of milk contamination by hygienic indicator, the frequency of testing cheese batches, and the number of sample units of cheese tested.

In summary, this work involves three major components as depicted in Figure 1.3. Using the foundational principles of Quantitative Microbial Risk Assessment (QMRA), we develop a QMRA model for simulating the contamination by various pathogens of raw milk soft cheese. The primary focus is on incorporating the effects of different pathogens into one QMRA model and assessing their combined effect on consumer health using suitable metrics,

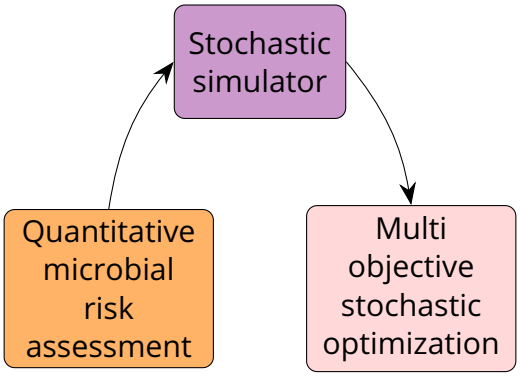


Figure 1.3: Proposed workflow in the context of this thesis.

such as Disability Adjusted Life Years (DALY) (Cassini et al., 2018). This model is implemented as a stochastic simulator used to simulate several batches, allowing us to estimate the ultimate quantities of interest using a simple Monte Carlo estimator. Given that having a substantial computational simulation cost is a major bottleneck for such estimators, our emphasis is on optimizing the implementation of this simulator. This leads us to dedicate a segment of our research to exploring the possibility of leveraging the intrinsic properties of the simulator to reduce computational time. More precisely, we aim to develop adaptive integration methods with better convergence rates (see, e.g., Basak et al., 2022b, Novak, 1992) than simple Monte Carlo, focusing on a special class of problems with properties like monotonicity and boundedness.

Once the QMRA simulator is implemented, it can be formalized as a stochastic simulator, whose outputs should be minimized. To achieve this goal, we use the framework of Bayesian optimization for multiobjective stochastic simulators. The objective is to build an algorithm that is easy to implement and not as expensive compared to other Bayesian approaches which are based on computationally-intensive criteria for point selection.

1.3 . Outline of the Manuscript and Contributions

This manuscript is organized into three main chapters.

Chapter 2 introduces the multi-pathogen quantitative risk assessment model, detailing its implementation, applicability, and user perspectives. To our knowledge, this represents the first proposition in microbiological risk assessment literature of a QMRA model incorporating the effects of three distinct pathogens present in a food item. In this chapter, Section 2.2 elucidates the components and methodological underpinnings of the QMRA model. Section 2.3 delves into the practical aspects and implementation of the model, encompassing data descriptions, parameter choices, and the employed mathemat-

ical/statistical techniques. Section 2.4 presents numerical experiments illustrating the model's functionalities, emphasizing the impact of intervention parameters on model outputs. The chapter concludes with a discussion on the model's limitations, applicability, and potential avenues for future development.

Chapter 3 is about numerical integration techniques, exploring their potential to enhance the cost-efficiency of the QMRA model's implementation. This chapter investigates both nonsequential and sequential methods, detailed respectively in Section 3.2 and Section 3.3. The contributions here are twofold: a formal literature review on available methods for this specific integration problem, and proposed extensions to augment their practical applicability within a fixed cardinality framework. Section 3.4 outlines numerical experiments conducted to test these proposed enhancements, and Section 3.5 demonstrates an application case on the QMRA simulator. Finally, Section 3.6 discusses the challenges, limitations, and prospective improvements related to the QMRA model's application.

Chapter 4 outlines the Bayesian framework for the multiobjective optimization of computationally intensive stochastic simulators, introducing an algorithm for efficient estimation of Pareto optimal solutions. This chapter revisits two established approaches to devise a sampling criterion for optimization algorithms—"maximal uncertainty sampling" and "stepwise uncertainty reduction"—each detailed in Section 4.3 and Section 4.4 respectively. The contributions of this chapter are threefold: introduction of a new algorithm for multiobjective stochastic optimization (MOSO) focused on estimating the Pareto front, a numerical benchmark comparing the proposed algorithm to existing ones, and an extension of the PALS (Barracosa et al., 2021) algorithm specifically tailored to the QMRA model. The effectiveness of this algorithm is illustrated through a case study applying it to the multi-pathogen QMRA model to estimate Pareto optimal solutions. The chapter concludes with discussions on the proposed algorithms in relation to their performance in estimating the Pareto front and set, accompanied by insights into the constraints and anticipated enhancements of each approach.

1.4 . Communications

Chapter 2 is an article in preparation with Laurent Guillier, Julien Bect, Janushan Christy, Fanny Tenenhaus-Aziza and Emmanuel Vazquez. An oral presentation of this work was made by Basak et al. (2023b), which is an extension of the work Basak et al. (2024). Chapter 3 is the extension of the work Basak et al. (2022b). Chapter 4 is based on the work presented in the communications Basak et al. (2022a, 2023a).

Journal publication

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: $10.3897/\mathrm{fmj}.5.109502$. URL https://doi.org/10.3897/fmj.5.109502

Communications with proceedings

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. 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, S. Petit, J. Bect, and E. Vazquez. Numerical issues in maximum likelihood parameter estimation for gaussian process interpolation. In *Machine Learning, Optimization, and Data Science*, pages 116–131. Springer International Publishing, 2022c. ISBN 978-3-030-95470-3

Communications without proceedings

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. Bayesian multi-objective optimization for stochastic simulators. In *MASCOT-NUM 2023*, 2023a

2 - Quantitative risk assessment model

2.1 . Introduction

Microbiological food safety is a major challenge for the food sector (see, e.g., Plaza-Rodriguez et al., 2018). In this context, the microbiological food safety community—including food authorities, food industries, and food research institutes—have invested research efforts into the field of Quantitative Microbial Risk Assessment (QMRA). The aim is to establish risk-based control measures (see, e.g., Koutsoumanis and Aspridou, 2016). QMRA is a part of microbial risk analysis, which involves risk assessment, risk management, and risk communication (World Health Organization, 1997).

Microbial risk assessment enables the evaluation of the likelihood of illness caused by pathogenic microorganisms and environmental factors that impact microbial growth. According to Codex Alimentarius Commission (1999), the framework for executing a QMRA for pathogens is built on several foundational components: hazard identification, exposure assessment, hazard characterization, risk characterization, and risk management options.

Hazard identification entails recognizing microbiological agents—such as bacteria, pathogens, and viruses—that exist in food and may lead to adverse health effects. This identification process is initiated after the problem formulation (Ungaretti Haberbeck et al., 2018). It involves delineating a list of microbial pathogens associated with the risk assessment in question.

Exposure assessment identifies and characterizes the pathways of the exposure to the microbial hazards and quantifies the exposure levels to estimate the magnitude of intake via consumption of a given food commodity.

Hazard characterization then converts these levels of exposure into a probability, representing the likelihood of adverse health effects associated with the hazard.

Risk characterization consolidates this information, offering a quantitative estimation of the likelihood, along with its associated uncertainties, of experiencing known or potential adverse health effects in a specific population. This estimation draws on data from hazard identification, hazard characterization, and exposure assessment (Codex Alimentarius Commission, 1999).

Generally, cheeses are considered safe and nutritious food, but food-borne illnesses related to cheese consumption occur (see, e.g., Dubois-Brissonnet et al., 2022). In the microbial risk assessment literature, there exist a number of QMRA studies on the contamination of raw milk soft cheese (see, e.g., Campagnollo et al., 2018, Lindqvist et al., 2002, Ramos et al., 2021, Sanaa et al., 2004, Tenenhaus-Aziza et al., 2014), where the authors have proposed methods not only to compute bacterial prevalence and contamination at the time

of consumption but also to identify major parameters contributing to the risk, using simulation studies (usually through "what-if" scenarios).

Among food-borne pathogens, Shiga-toxin producing *Escherichia coli* (STEC) in soft cheese is a notable concern (Farrokh et al., 2013). The definition of the virulence potential of STEC is not straightforward (Lindqvist et al., 2023). Depending on the strains' characteristics and the exposed population, the symptoms can range from mild to severe illnesses, such as Haemolytic and Uremic Syndrome (HUS), which is a leading cause of renal failure in young children. The growth and survival of pathogenic STEC serotypes throughout different phases of cheese production, were studied by Maher et al. (2001) and Miszczycha et al. (2016). In this context, Perrin et al. (2014) proposed a stochastic QMRA model to assess the risk of HUS associated with the five Main Pathogenic Stereotypes of STEC (MPS-STEC) in raw milk soft cheeses, and explored the role of control measures for minimizing the risk of illness.

Building on the work of Perrin et al. (2014), we introduce a new farm-to-fork QMRA multipathogen model, integrating the impacts of three bacteria—specifically, STEC, non-typhoidal *Salmonella*, and *Listeria monocytogenes* which can be potentially present in raw milk (see, e.g., Costanzo et al., 2020, D. et al., 2009, Sanaa et al., 2004). Like STEC, *Salmonella* and *Listeria monocytogenes* are hazards that, upon infecting humans, can trigger distinct health conditions. *Salmonella* can lead to a condition called salmonellosis, which is characterized by symptoms such as diarrhea, abdominal pain, fever, and vomiting. *Listeria monocytogenes*, on the other hand, can cause listeriosis, a more severe illness (see, e.g., Camargo et al., 2017, Leclercq et al., 2014), that can lead to fever, muscle aches, nausea, diarrhea, and in severe cases, can even result in meningitis or septicemia.

This new multipathogen QMRA model integrates the pre-harvest and post-harvest intervention steps in the cheese production processes, which are used as control measures. This integration will enable the evaluation of their impact on the risk of illness.

This chapter is structured as follows. Section 2.2 describes the model with the underlying assumptions and its components, along with their functionalities, developed according to the World Health Organization et al. (2021) framework. Section 2.3 elaborates on the implementation of the model, detailing the selection of model parameters and the mathematical and statistical techniques employed. Section 2.4 presents numerical results regarding bacterial prevalence and risk reduction, derived from exploring various intervention scenarios. Finally, Section 2.5 provides a discussion on the model's usage, applicability, and perspectives for future work.

2.2 . Description of the proposed hierarchical model

2.2.1 . Model overview

The multipathogen QMRA model builds upon the QMRA model for STEC proposed by Perrin et al. (2014) and the R implementation provided by Basak et al. (2024).

The QMRA model is used to assess the risk of microbial contamination by examining the evolution of pathogens throughout the entire cheese-making process, from the farm where milk is produced to the consumer's fork where it is consumed. This farm-to-fork model can be regarded as a stochastic simulator, a computational model used for simulating a complex system that incorporates inherent randomness. The model consists of two hierarchical components: a batch-level simulator and an output module. The batch-level simulator models all the various steps of the cheese manufacturing process, beginning with the collection of milk from a specific number of farms on a given day and continuing through the production process. The production process usually lasts up to 14 days until the cheese ripening step, followed by a cheese storage step until the 22nd day, after which the produced batch of cheese is sent to the market (see, e.g., Perrin et al., 2014). A typical batch of cheese usually contains $22,000-23,000$ cheeses of 250 g, produced using a total volume of $50,000$ L of raw milk, though it can vary depending on the cheese producer. The outputs of interest corresponding to a particular cheese batch is produced by the batch-level simulator, which are then used by the output module, that produces an estimate for the impact of the food-borne illnesses and the intervention costs.

Notations and abbreviations Throughout this chapter, we adopt a notation convention in which, depending on the variables used for the indices, it is implicitly assumed that these variables have values in different sets of pathogens, as summarized in Table 2.1.

Table 2.1: Notations and corresponding pathogen classification.

Index notation	Pathogen membership set
x	MPS-STEC, *Salmonella*, *Lm*
$x^{\dagger}$	MPS O157:H7, MPS nonO157:H7, HV *Salmonella*, LV *Salmonella*, *Lm*
$x^{\ddagger}$	MPS O157:H7, MPS nonO157:H7, *Salmonella*, *Lm*
$x^{\dagger\dagger}$	MPS-STEC, HV *Salmonella*, LV *Salmonella*

Here, MPS-STEC denotes the Main Pathogenic Serotypes of Shiga Toxin Producing *Escherichia coli*, and it is classified into two subclasses based on serotype markers, namely O157:H7 and non O157:H7 (Perrin et al., 2014). For *Salmonella*, we consider two subclasses corresponding to their high and low virulent strains, respectively abbreviated by HV *Salmonella* and LV *Salmonella*. In mathematical notations, these are further compressed as HV-*Salmo* and LV-*Salmo*. For *Listeria monocytogenes*, the abbreviation *Lm* is used.

Modules The batch-level simulator has four modules: a farm module followed by a preharvest intervention step, a cheese production module, a consumer module, and a postharvest sampling module.

The sets of input parameters, as detailed in Section 2.3.3, for each of the modules, are aggregated in a vector denoted by $\theta = \{\theta^{\text{farm}}, \theta^{\text{cheese}}, \theta^{\text{con}}, \theta^{\text{post}}\}$, which forms the input parameter of the simulator.

The farm module, as explained in Section 2.2.3.1, models the collection of milk from different farms and outputs the initial pathogen concentration Y_x^{milk} in the aggregated milk tank, for the pathogen x (see Table 2.1). (Note that we use capital letters in the description of the model to designate random variables. This will be made explicit in the next sections.) The farm module also implements the preharvest intervention step, as explained in Section 2.2.5.1, which prevents contaminated farm milk from entering the production process.

The cheese module, as explained in Section 2.2.3.2, models the evolution of pathogens during the cheese production process and simulates the colony size $Y_{x^\ddagger}$ and the average number of colonies $\lambda_{x^\dagger}^{\text{colony}}$ in a single cheese from that specific batch. Note that $Y_{x^\ddagger}$ and $\lambda_{x^\dagger}^{\text{colony}}$ are random variables that depend on Y_x^{milk} and θ^{cheese}.

The consumer module, as described in Section 2.2.4.1, models the impact of the ingested dose of pathogen by the consumer and estimates the risk of illness.

The postharvest intervention step, as described in Section 2.2.5.2, implements the microbial cheese sampling plan and estimates the probability of detecting contamination in the produced batch of cheese.

Given the input vector θ, the batch-level simulator yields the following outputs relevant to a particular batch: the milk loss per batch M^{batch} due to the preharvest intervention step, the probability of rejecting the batch of cheese P^{batch} due to the postharvest intervention step, and the batch risk R_x^{batch}, that is, the risk of encountering food-borne illnesses due to pathogen x, if a portion of 25 g of cheese is consumed from that specific batch of cheese. Again, note that M^{batch}, P^{batch} and R_x^{batch} are random variables in what follows. When the batch level simulator is run once, it will simulate a sample value for these random variables, conditional on θ.

Figure 2.1 offers a schematic diagram of the batch-level simulator, encompassing several modules. The output module, as explained in Section 2.2.6.1, is used to simulate several batches of cheese, that is, several sample values of M^{batch}, P^{batch} and R_x^{batch}, and to estimate the final quantities of interest, e.g., the prevailing risk of illness, average milk loss, and average probability of rejecting a batch of cheese.

2.2.2 . Hazard identification

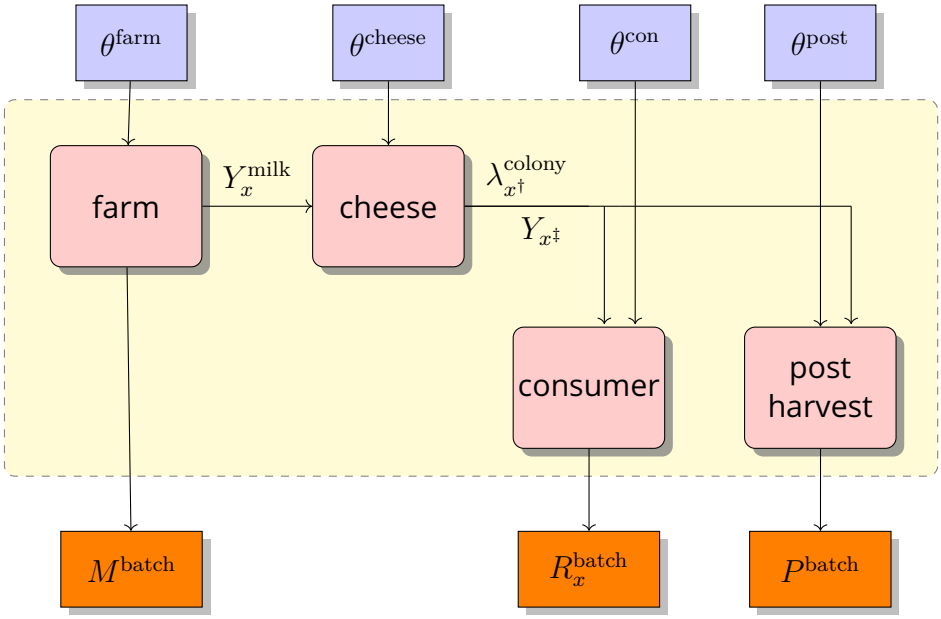


Figure 2.1: Batch level simulator: The four modules of the simulator, namely, farm, cheese, consumer and postharvest, are shown in pink boxes, along with their corresponding set of inputs in violet boxes. The outputs corresponding to a simulation of a single, namely batch are shown using orange boxes.

The three pathogens identified as hazards in this study—the main pathogenic serotypes of Shiga-toxin producing *Escherichia coli* (also known as MPS-STEC), *Salmonella*, and *Lm*—can survive or grow during cheese making, particularly in raw milk soft cheeses (see, e.g., Costanzo et al., 2020). These bacterias present in the intestines of lactating dairy animals, and can be transmitted through fecal matter to their udders, thereby contaminating milk during the milking process (Gopal et al., 2015).

Main pathogenic serotypes of STEC. The five main pathogenic serotypes of STEC (MPS-STEC) identified thus far in Europe are O157:H7, O26:H11, O103:H2, O111:H8, and O145:H28. According to EFSA (2017), EFSA BIOHAZ Panel et al. (2020), 14 outbreaks involving STEC in milk, dairy, and cheese products were recorded between 2012 and 2017, affecting 775 individuals, with the primary causative agents being Shiga toxin-producing *E. coli* (STEC). In 2005, in France, an outbreak of Hemolytic Uremic Syndrome (HUS) was reported, linked to the contamination of raw milk soft cheese with *E. coli* O26 and O80, followed by another outbreak in 2009 among children up to 15 years (King et al., 2009) and in 2019 (Jones et al., 2019). However, other pathogenic serotypes have caused major outbreaks. In 2010, atypical STEC serotypes O104:H4 caused a large outbreak in Germany (see, e.g., EFSA, 2012, Frank et al., 2011). Recently, the O80 serotype has arisen in Europe (see, e.g., Bruyand et al., 2019).

Listeria monocytogenes. The presence of *Lm* in raw milk and cheese has been extensively reported (see, e.g., Dalzini et al., 2016), with its widespread occurrence and potential for contamination possible at any stage of the production chain. Due to cold tolerance, i.e., the ability to grow at refrigeration temperatures as low as -1.5 °C, and its capacity to form resilient biofilms resistant to sanitation, *Lm* exhibits the ability to persist and survive in various environments (see, e.g., McIntyre et al., 2015). According to European Food Safety Authority and for Disease Prevention & Control (2022), there were 2183 confirmed invasive human cases of listeriosis in 2021. Cheese was estimated to be the origin of 1% of the listeriosis cases in Europe according to risk assessment models (Ricci et al., 2018).

Salmonella. Salmonellosis is recognized as one of the most common bacterial food-borne illnesses in humans, with several outbreaks reported from *Salmonella* contamination in raw milk cheese (see, e.g., Robinson et al., 2020, Ung et al., 2019). *Salmonella* is commonly found in the intestinal tract of lactating animals, and milk contamination primarily occurs during the milking process (Ruzante et al., 2010). Among various serotypes, *Salmonella* Dublin, *Salmonella* Newport, and *Salmonella* Typhimurium are commonly linked to salmonellosis in both calves and adult cows, leading to varying degrees of illness. Additionally, dairy animals have been found to carry *Salmonella* serotypes such as *Salmonella* Cerro, *Salmonella* Kentucky, *Salmonella* Mbandaka, and *Salmonella* Montevideo without showing any symptoms, while shedding the bacteria in their feces (see, e.g., Bonifait et al., 2021, Van Kessel et al., 2012).

2.2.3 . Exposure Assessment

2.2.3.1 Farm module

At the farm level, milk is collected from a mixed herd of lactating animals, which can potentially includes both infected and non-infected cows, and stored in a bulk tank. During the milking process, pathogens potentially present in the fecal matter of infected cows can be transmitted into the bulk tank through their contaminated udders. The Bulk Tank Milk (BTM) is then collected from different farms and mixed into an Aggregated Milk Tank (ATM), which is used for cheese production. In the QMRA model, the farm module replicates this entire scenario of milk collection and simulates the concentration in CFU/ml (Colony forming units per mili liters) of all the pathogens in the ATM. The input parameters of the farm module are listed in Table 2.5.

The concentration $Y_{\mathrm{Lm}}^{\mathrm{milk}}$ of *Lm* in the ATM was studied by Sanaa et al. (2004). Following this work, we model $Y_{\mathrm{Lm}}^{\mathrm{milk}}$ as a log-normal random variable:

$$Y_{\mathrm{Lm}}^{\mathrm{milk}} \mid \theta^{\mathrm{farm}} \sim \mathrm{Lognormal}(\mu_{\mathrm{Lm}}, \tau_{\mathrm{Lm}})\,. \tag{2.1}$$

Due to very low contamination level, insufficient knowledge and the un-

availability of a reliable method to directly determine the pathogen concentration in milk for STEC and *Salmonella*, we employ the indirect approach of estimating concentration used by Perrin et al. (2014) for the QMRA model for STEC. This approach relies on collected data on *E. coli* concentration in farm milk, assuming that *E. coli* and other pathogen strains follow the same fecal routes.

Suppose that, for each farm indexed by $i = 1, 2, \ldots, N^{\text{farms}}$, the milk is collected into a BTM from N_i^{cow} cows; then the concentration of pathogen $x^{\dagger\dagger}$ in the BTM corresponding to farm i is denoted by $Y_{x^{\dagger\dagger},i}^{\text{milk}}$ and is obtained as,

$$Y_{x^{\dagger\dagger},i}^{\text{milk}} = Y_i^{\text{EC}} \cdot \frac{\overline{F_i^{x^{\dagger\dagger}}}}{\overline{F_i^{\text{EC}}}}, \tag{2.2}$$

where Y_i^{EC} denotes the concentration (CFU/mL) of *E. coli* in BTM, $\overline{F_i^{x^{\dagger\dagger}}}$ and $\overline{F_i^{\text{EC}}}$ respectively denote the average concentration (CFU/gram) in the fecal matter for the $x^{\dagger\dagger}$ pathogen and *E. coli*, coming from all the cows.

To obtain Y_i^{EC}, $\overline{F_i^{x^{\dagger\dagger}}}$ and $\overline{F_i^{\text{EC}}}$, the module first models the number of infected cows using a binomial distribution $k_i^j \mid \theta^{\text{farm}} \sim \text{Binomial}(N_i^{\text{cow}}, p_j)$, in the i-th farm, with the corresponding class probabilities p_j of the two major pathogen classes $j \in \{\text{STEC, } \textit{Salmonella}\}$. For STEC, the proportion p_{STEC} of infected cows is estimated using a logit model (following Perrin et al., 2014). Next, the number of cows affected by the subclass serotypes is modeled, using the respective class probabilities:

$$k_i^{\text{MPS-STEC}} \mid \theta^{\text{farm}} \sim \text{Binomial}(k_i^{\text{STEC}}, p_{\text{MPS-STEC}}), \tag{2.3}$$

$$k_i^{\text{HV-Salmo}} \mid \theta^{\text{farm}} \sim \text{Binomial}(k_i^{\text{Salmo}}, p_{\text{HV-Salmo}}). \tag{2.4}$$

The number of *Salmonella*-infected cows carrying the low virulent serotype, is computed as $k_i^{\text{LV-Salmo}} = k_i^{\text{Salmo}} - k_i^{\text{HV-Salmo}}$. For the j-th infected cow, $1 \leq j \leq k_i^{x^{\dagger\dagger}}$, in the i-th farm, $1 \leq i \leq N^{\text{farm}}$, the concentration in the fecal matter is modeled according to a Weibull distribution for MPS-STEC (Perrin et al., 2014),

$$F_{i,j}^{\text{MPS-STEC}} \mid \theta^{\text{farm}} \sim \text{Weibull}(a^{\text{weibull}}, b^{\text{weibull}}),$$

and the concentrations of the two serotypes HV-*Salmonella*, LV-*Salmonella* are modeled using a log-normal distribution (Bonifait et al., 2021),

$$\log_{10}(F_{i,j}^{\text{Salmo}}) \mid \theta^{\text{farm}} \sim \mathcal{N}(a^{\text{Salmo}}, b^{\text{Salmo}}).$$

For each farm, the average concentration of pathogen $x^{\dagger\dagger}$ is

$$\overline{F_i^{x^{\dagger\dagger}}} = \frac{1}{N_i^{\text{cow}}} \sum_{j=1}^{k_i^{x^{\dagger\dagger}}} F_{i,j}^{x^{\dagger\dagger}}.$$

The concentration of *E. coli* (CFU/mL) in a BTM Y_i^{EC}, is modeled by a log-normal distribution:

$$Y_i^{\mathrm{EC}} \mid \theta^{\mathrm{farm}} \sim \mathrm{Lognormal}(\alpha_i, \sigma_i). \tag{2.5}$$

The average $\overline{F_i^{\mathrm{EC}}}$ of individual *E. coli* concentrations in fecal matter for each cow, denoted by $F_{i,j}^{\mathrm{EC}}$, $j = 1, 2, \ldots, N_i^{\mathrm{cow}}$, is obtained using the model

$$\begin{aligned} &\log 10(F_{i,j}^{\mathrm{EC}}) \mid \theta^{\mathrm{farm}} \sim \mathcal{N}(\mu^{\mathrm{ecoli}}, \tau^{\mathrm{ecoli}}), \\ &\overline{F_i^{\mathrm{EC}}} = \frac{1}{N_i^{\mathrm{cow}}} \sum_{j=1}^{N_i^{\mathrm{cow}}} F_{i,j}^{\mathrm{EC}}. \end{aligned} \tag{2.6}$$

Each of the BTMs is tested (also known as milk sorting, see Section 2.2.5.1) for *E. coli* concentration and accordingly accepted or rejected for cheese production. Let S denote the set of farms that qualify after milk sorting and let $N^{\mathrm{farms,sorted}} = |S|$. After milk sorting, milk from all the qualified BTMs is collected into a single ATM. The final concentration (CFU/mL) of pathogen $Y_{x^{\dagger\dagger}}^{\mathrm{milk}}$ in this ATM can be written as

$$Y_{x^{\dagger\dagger}}^{\mathrm{milk}} = \sum_{i=1}^{N^{\mathrm{farms}}} \left(Y_{x^{\dagger\dagger},i}^{\mathrm{milk}} \cdot \frac{V_i \mathbb{1}_{\{i \in S\}}}{\sum_{i=1}^{N^{\mathrm{farms}}} V_i \mathbb{1}_{\{i \in S\}}} \right), \tag{2.7}$$

where V_i is the volume of milk in liters produced by the ith farm.

Module outputs. The farm module models the concentrations (CFU/mL) of four different pathogenic serotypes in milk in the ATM, namely, MPS-STEC, HV *Salmonella*, LV *Salmonella* and *Lm*. It also yields the milk loss M^{batch} (in Liters), due to the preharvest milk testing step, associated with the production of that particular batch of cheese. In addition it models the number of farms discarded due to milk testing $N^{\mathrm{farms}} - N^{\mathrm{farms,sorted}}$, and the total volume of milk put in production, or in other words the volume of milk in the ATM $\sum_{i=1}^{N^{\mathrm{farms}}} V_i \mathbb{1}_{\{i \in S\}}$.

2.2.3.2 Cheese module

The inputs for the cheese module are the initial pathogen concentrations from the farm module and a set of parameters denoted by θ^{cheese} detailed in Table 2.10.

After being collected in the ATM, the milk undergoes several processing steps, including milk storage, molding, draining, salting, ripening, and cheese storage. These steps can be categorized into two phases, namely the liquid phase and the solid phase, respectively illustrated by two schematic diagrams in Figure 2.2 and Figure 2.3.

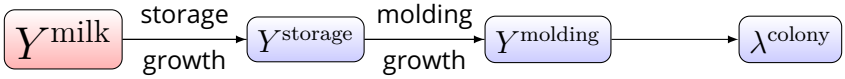


Figure 2.2: Liquid phase steps: storage and molding, models the pathogen concentration, starting from initial concentration Y^{milk}.

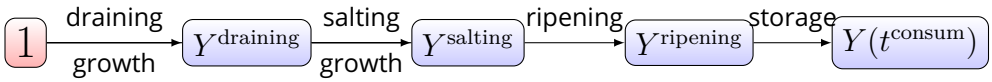


Figure 2.3: Solid phase steps: draining, salting, ripening and cheese storage, models the evolution of colony size, starting from 1 bacteria.

During the pre-molding steps in the liquid phase, a growth in the bacteria concentration is observed, for all the three pathogens. At the end of the liquid phase, the pathogen cells are presumed to become immobilized within the cheese matrix, resulting in the formation of colonies. At this stage, the cheese module estimates the average number of colonies formed in a single cheese. After entering the solid phase, the cheese module models the evolution of the colonies starting from one single colony. Finally it estimates the colony sizes for different pathogens at the time of consumption. The rest of this section is organized in three different parts, corresponding to the evolution of pathogens during the different cheese processing steps.

Growth phase. For all three pathogens, the growth phase occurs until the salting step. Starting from the farm module outputs, the cheese module models the growth of the pathogens over the different steps using an ordinary differential equation

$$\begin{cases} \frac{\mathrm{d}Y_x}{\mathrm{d}t} = \mu_x^{\max}(t) \cdot Y_x(t) \cdot \left(1 - \frac{Y_x(t)}{y^{\max}}\right), \\ Y_x(0) = Y_x^{\text{milk}} \end{cases} \tag{2.8}$$

where $\mu_x^{\max}(t)$ stands for the maximum growth rate (in h^{-1}) and $y^{\max}$ is a parameter that represents the hypothetical maximum population of pathogen strains in milk or cheese. The maximum growth rate $\mu_x^{\max}(t)$, as shown in Figure 2.5a, is modeled against time t according to Augustin et al. (2005) using the optimal growth rate parameter μ_x^{opt}, several physico-chemical parameters $\{d, \mathrm{pH}, T, \mathrm{aw}\}$, and their nominal values, collectively denoted as $\theta(\mu^{\max})$ and listed in Table 2.6. Figure 2.4 shows the variation in physico-chemical parameters over time, as studied by Perrin et al. (2014).

Remark 1. For *Salmonella*, the overall initial concentration $Y_{\text{Salmo}}^{\text{milk}}$ is computed as the sum of the concentrations of its two subclass serotypes.

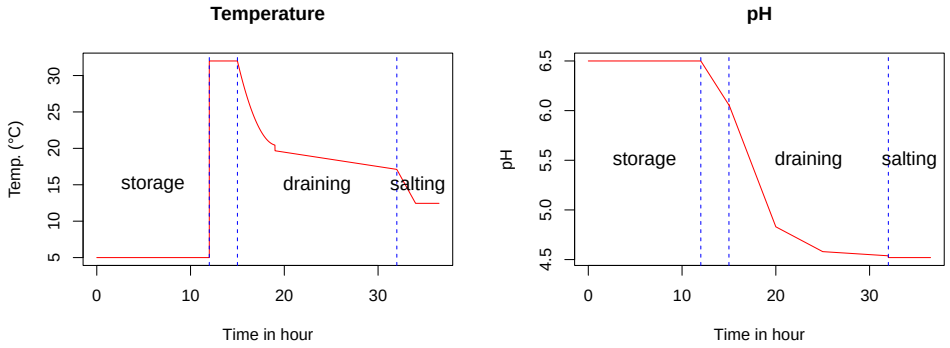


Figure 2.4: Dynamic physico-chemical parameters, namely pH and temperature, during milk storage, molding, draining and salting steps of cheese production, separated by blue dotted lines.

The concentrations during the liquid phase, specifically at the end of the storage step Y_x^{storage} and molding step Y_x^{molding}, are computed using (2.8), as depicted in Figure 2.5b.

Corresponding to the three major pathogens, we consider a total of five different subclass serotypes denoted by $x^\dagger$.

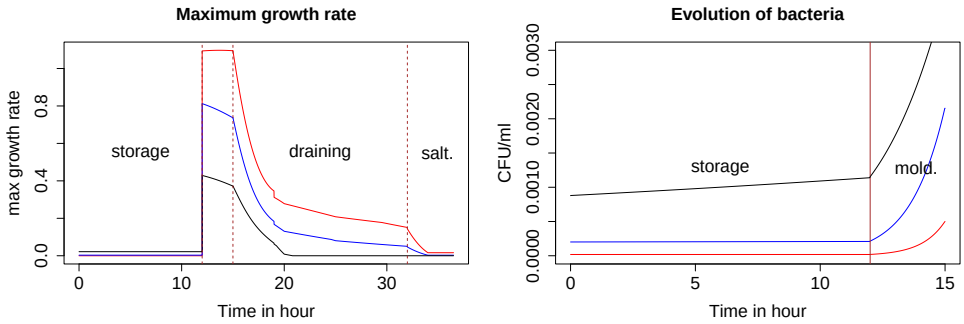
At the end of the liquid phase, conditional on the concentration of the molding step Y_x^{molding}, the number of colonies for each pathogen subclass serotypes are modeled as a Poisson variable $N_{x^\dagger}^{\text{colony}} \mid Y_x^{\text{molding}} \sim \text{Poisson}(\lambda_{x^\dagger}^{\text{colony}})$, with the mean computed as

$$\lambda_{x^\dagger}^{\text{colony}} = Y_x^{\text{molding}} \cdot v^{\text{cheese}} \cdot w^{\text{loss}} \cdot p_{x^\dagger}, \tag{2.9}$$

where for MPS-STEC and *Salmonella*, $p_{x^\dagger}$ denotes the class probability of the corresponding subclass serotype, whereas $p_{\text{Lm}} = 1$ for *Lm*. The parameters v^{cheese} and w^{loss} denote the amount of milk used for a single cheese and the proportion of water loss during the molding step, respectively.

Starting from the draining step, with an initial size of 1 CFU, (2.8) models growth of the size of that single colony forming unit, until the salting step. The evolution of each colony inside a particular cheese (of mass $250\,\text{g}$) is assumed to be identical during the growth phase due to the same environmental conditions.

Decline phase. At the end of the salting step, when the batch of cheese enters the ripening step, there is a decline in the population of the colonies. The ripening step lasts until the 14-th day of the production (Perrin et al., 2014), followed by a cheese storage step until the 22-nd day (Basak et al., 2024), and after that the cheese is sent to the market. The cheese is consumed on the t^{consum}-th day, which is a triangular-distributed random variable with minimum possible value 22.



(a) Maximum growth rate (b) Growth during storage and molding

Figure 2.5: The three pathogens STEC, *Salmonella* and *Lm* are shown in red, blue and black lines respectively. The evolution of bacteria is calculated starting from average values of initial concentrations, for a baseline scenario with no preharvest intervention step.

For MPS-STEC and *Salmonella* this decline phase is considered as a continuous process which lasts until the cheese is consumed, whereas for *Lm*, only a decline in the colony population is considered. Let t^{salting} be the time taken (in hours) until the salting step, then for any time point $t > t^{\text{salting}}$, the cheese module computes four reference colony sizes $Y_{x^{\ddagger}}(t)$, for the four pathogenic serotypes denoted by $x^{\ddagger}$. These reference colony sizes represent the size of the respective pathogen colonies inside a particular cheese in the batch, without taking into account the inter-cheese variability. Here, we assume that all colonies within a particular cheese have the same decline phase, indicating no intra-cheese variability.

Following Perrin et al. (2014), the decline phase for MPS-STEC is modeled differently for its different subclass serotypes, indexed by $s \in \{\text{MPS O157:H7}, \text{MPS nonO157:H7}\}$. For MPS-STEC the reference colony size at time point t (in days), is computed as $Y_s(t) = Y^{\text{salting}}_{\text{MPS-STEC}} \cdot 10^{-\rho_s \cdot (t - t^{\text{salting}}/24)}$, with the subclass serotypes distinguished by the decline rate parameter ρ_s.

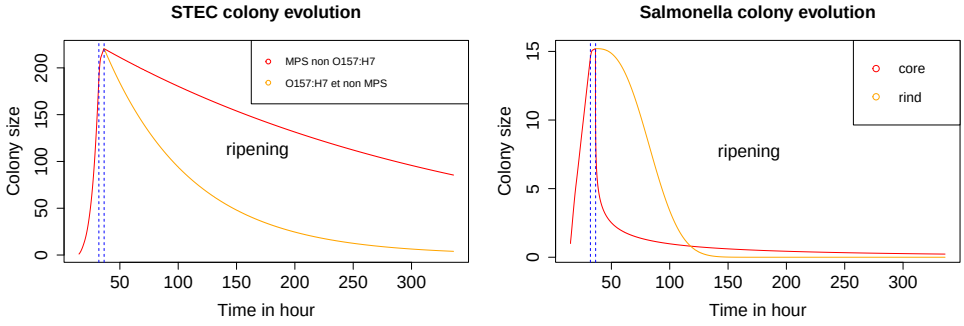
The decline phase for *Salmonella* is dependent on the position of the colonies in the cheese matrix. The decline is modeled using different set of parameters for the core region colonies and the rind region colonies (Gonzales-Barron et al., 2022), namely, $\{\delta_s, p_s\}$, for $s \in \{\text{core}, \text{rind}\}$. The reference colony size at time t, for serotype s is computed as $Y_s(t) = Y^{\text{salting}}_{\text{Salmo}} \cdot 10^{-((t - t^{\text{salting}}/24)/\delta_s)^{p_s}}$. The reference colony size $Y_{\text{Salmo}}(t)$ for *Salmonella* is obtained by averaging over the core and rind region with respect to the proportion of colonies in those region $p_{\text{core.rind}}$.

Following Sanaa et al. (2004), for *Lm*, there is a decrease in the population after the salting step during curd acidification in cheese vats and molds, and it is modeled using log apparent kills, $Y^{\text{post.salting}}_{\text{Lm}} = Y^{\text{salting}}_{\text{Lm}} \cdot 10^{-\rho_{\text{Lm}}}$, with a

triangular distributed decline parameter ρ_{Lm}.

Second growth phase. The second growth phase is only observed for *Lm* (Sanaa et al., 2004), and it initiates depending on the position of the colonies in the core and rind region. Challenge test data collected under the Arti-SaneFood project were used identify the favorable environmental conditions, namely, the physico-chemical parameters pH_{Lm}, T_{Lm} that initiate the second growth step. These parameters were used to estimate the maximum growth rate for *Lm*, and the colony size for the core and rind regions at time t was modeled separately using (2.8), starting from an initial concentration $Y_{\mathrm{Lm}}^{\mathrm{post.salting}}$. It was observed that colonies in the core region (comprising 90% of the total colonies) enter the second growth phase after 20 days of the salting step, while the colonies at the surface region (making up 10% of the total colonies) initiate their second growth phase after 7 days of the end of salting step, or equivalently, the start of the ripening step. The reference colony size $Y_{\mathrm{Lm}}(t)$ is obtained by averaging over the core and rind regions, similarly as *Salmonella*.

The evolution of colonies for the three pathogens, during the solid phase, is depicted in Figure 2.6a, Figure 2.6b, and Figure 2.7.



(a) STEC serotypes (b) Core and rind colonies of *Salmonella*

Figure 2.6: Evolution of colony sizes starting from unit size, for MPS-STEC and *Salmonella*, during the draining, salting and ripening steps, separated by blue dotted lines.

Module outputs. For all the five subclass serotypes, the expected number of colonies $\lambda_{x^\dagger}^{\mathrm{colony}}$ at a particular time point t is adjusted with respect to the corresponding reference colony size of the subclass serotype (for example, $Y_{\mathrm{O157}}(t)$ and $Y_{\overline{\mathrm{O157}}}(t)$ for MPS-STEC) or the reference colony size of the pathogen itself (for example, $Y_{\mathrm{Salmo}}(t)$ and $Y_{\mathrm{Lm}}(t)$ for *Salmonella* and *Lm* respectively), at time t. If the reference colony size $Y_{x^\ddagger}(t)$ falls below 1, we

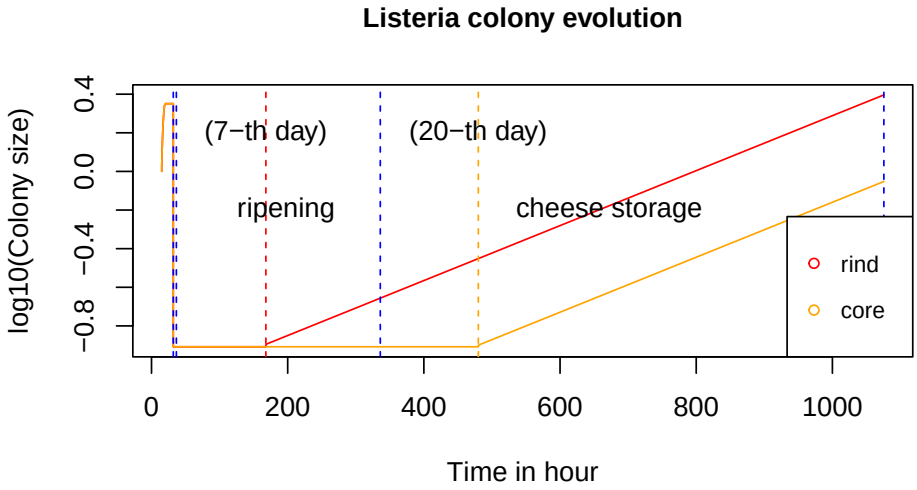


Figure 2.7: Evolution of colony sizes starting from unit size, for *Lm*, during the draining, salting, ripening and cheese storage step, separated by blue dotted lines. The red and orange dotted lines respectively indicates the starting time of the second growth step, corresponding to the rind and core region colonies.

choose the assumption that the colonies have disappeared with a probability $Y_{x^\ddagger}(t)$, and the corresponding adjusted expected number of colonies is obtained as $\lambda_{x^\dagger}^{\text{colony}}(t) = \lambda_{x^\dagger}^{\text{colony}} \cdot Y_{x^\ddagger}(t)$. If $Y_{x^\ddagger}(t) > 1$, the expected number of colonies remains unchanged.

The outputs of the cheese module are the adjusted expected number of colonies $\lambda_{x^\dagger}^{\text{colony}}(t)$, for pathogen $x^\dagger$ and the corresponding reference size of the colonies $Y_{x^\ddagger}(t)$, for pathogen $x^\ddagger$. These quantities are computed at two specific time points, firstly, the time of testing samples from the cheese batches t^{test}, and secondly at the time of consumption t^{consum}.

2.2.4 . Hazard characterization

2.2.4.1 Consumer module

The consumer module describes the nature and probability of adverse human effects as a function of viable pathogen numbers ingested, termed the dose. The consumer module uses the outputs of the cheese module, that is, the adjusted expected number of colonies $\lambda_{x^\dagger}^{\text{colony}}(t^{\text{consum}})$ in a cheese and the reference size of one colony $Y_{x^\ddagger}(t^{\text{consum}})$, at the time of consumption t^{consum}. In addition to the cheese module outputs, the other inputs of the consumer module, denoted by θ^{con}, are listed in Table 2.8. The consumer module simulates the risk of the corresponding illness based on the cheese consumption

behavior of people in different age groups.

Inter-cheese variability. In the modeling of risk, considering inter-cheese variability is crucial as it accounts for the differences observed in the distribution of pathogenic colonies across different cheese samples. Specifically, even though the reference colony sizes $Y_{x^\ddagger}(t)$, as mentioned in Section 2.2.3.2, are defined without the inter-cheese variability, this variability is explicitly incorporated during risk computation.

For MPS-STEC, the final colony sizes for the two subclass serotypes, with the inclusion of the inter-cheese variability, follow a log-normal distribution (Perrin et al., 2014), described by $Y_s^{\text{colony}} = Y_s(t^{\text{consum}}) \cdot 10^{\epsilon_s}$, where $\epsilon_s \sim \mathcal{N}(\mu_{\epsilon_s}, \tau_{\epsilon_s})$ represents the parameter for inter-cheese variability, with s denoting the two subclass serotypes of MPS-STEC $s \in \{\text{MPS O157:H7, MPS nonO157:H7}\}$. While this assumption is made for MPS-STEC, we still posit that there is no intra-cheese variability, meaning all the colonies inside a specific cheese share an identical size Y_s^{colony} for serotype s (Perrin et al., 2014).

For *Salmonella*, the inter-cheese variability could be introduced through the variance of the decline step parameter δ_s, for $s \in \{\text{core, rind}\}$, denoted by τ_s^δ (Gonzales-Barron et al., 2022), as seen in Table 2.10. However, the current model assumes no inter-cheese variability for *Salmonella* and for *Lm* as well, resulting in their final colony size being equivalent to the reference colony size.

Model for the dose. The risk of getting the food-borne illness is dependent on the number of cells of pathogen $x^\dagger$ ingested by the consumer, which is referred to as the dose, denoted by $\Gamma_{x^\dagger}$. The dose is defined as the number of pathogenic cells present in a particular cheese serving of $25\,\text{g}$, which is obtained by multiplying the size of colonies $Y_{x^\ddagger}^{\text{colony}}$ with the number of colonies $N_{x^\dagger}^{\text{colony-serving}}$ in a cheese serving, distributed as

$$N_{x^\dagger}^{\text{colony-serving}} \mid \lambda_{x^\dagger}^{\text{colony}}, t^{\text{consum}} \sim \text{Poisson}\Big(\lambda_{x^\dagger}^{\text{colony}}(t^{\text{consum}}) \frac{w^{\text{serving}}}{w^{\text{cheese}}}\Big). \tag{2.10}$$

Probabilities of illness. To obtain probabilities of illness, three different dose-response models were used corresponding to the three different pathogens. Following Perrin et al. (2014), an exponential dose-response model was used for MPS-STEC, that uses the combined dose $\Gamma_{\text{MPS-STEC}} = N_{\text{O157}}^{\text{colony-serving}} \cdot Y_{\text{O157}}^{\text{colony}} + N_{\overline{\text{O157}}}^{\text{colony-serving}} \cdot Y_{\overline{\text{O157}}}^{\text{colony}}$, which is the sum of the doses corresponding to the two different serotype classes of MPS-STEC. The probability of getting the HUS disease by consuming a serving of $25\,\text{g}$ of cheese, conditional on $\text{Age} = \text{age}$ and dose $\Gamma_{\text{MPS-STEC}} = \gamma_{\text{MPS-STEC}}$, can be written as

$$P_{\text{HUS}}(\text{age},\ \gamma_{\text{MPS-STEC}}) = 1 - (1 - r_0 \cdot \exp(-k \cdot \text{age}))^{\gamma_{\text{MPS-STEC}}}. \tag{2.11}$$

For *Salmonella*, a Beta-Poisson dose-response model proposed by McCullough and Elsele (1951), World Health Organization (2002) is used, with model parameters (Strickland et al., 2023) dependent on the two subclass serotypes s, for $s \in \{\text{HV } \textit{Salmonella}, \text{LV } \textit{Salmonella}\}$,

$$P_{\text{Salmonellosis}}(\Gamma_s) = 1 - (1 + \frac{\Gamma_s}{\beta_s})^{-\alpha_s}. \tag{2.12}$$

Conditional on dose $\Gamma_s = N_s^{\text{colony-serving}} . Y_s^{\text{colony}}$, for each serotype s, the probability of getting salmonellosis from the consumption of a serving of $25\,\text{g}$ of cheese, is a weighted average, using the respective consumption probabilities based on expected number of high and low virulent colonies in a cheese serving:

$$\begin{aligned} &P_{\text{Salmonellosis}}(\Gamma_{\text{HV-Salmo}}, \Gamma_{\text{LV-Salmo}}) = \\ &\quad P_{\text{Salmonellosis}}(\Gamma_{\text{HV-Salmo}}) \cdot p^{\text{colony}}_{\text{HV-Salmo}} + P_{\text{Salmonellosis}}(\Gamma_{\text{LV-Salmo}}) \cdot p^{\text{colony}}_{\text{LV-Salmo}} . \end{aligned}$$

For *Lm*, the dose-response is 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). Conditional on dose $\Gamma_{\text{Lm}} = N_{\text{Lm}}^{\text{colony-serving}} \cdot Y_{\text{Lm}}^{\text{colony}}$, and population subgroup θ_g, the probability of getting listeriosis, from the consumption of a serving of $25\,\text{g}$ of cheese, is written as

$$P_{\text{Listeriosis}}(\theta_g, \Gamma_{\text{Lm}}) = \int_0^1 (1 - \exp(-r \cdot \Gamma_{\text{Lm}}))p(r; \theta_g)\,\text{dr}, \tag{2.13}$$

where r is the probability of developing listeriosis from the ingestion of a bacterial cell in a given, specific serving, and $p(r; \theta_g)$ represents the remaining individual (within-group) susceptibility variability and strain virulence variability in r.

Batch risk. The consumer module models the risk for each of the foodborne illnesses, associated with the consumption of a particular batch of cheese. For each pathogen x, the batch risk depends on a vector of stochastic internal variables of the QMRA model, denoted by Ξ_x, which characterizes a batch.

The vector Ξ_x includes the initial concentration Y_x^{milk} of the respective pathogen in the ATM, the stochastic parameters $\{d^{\text{storage}}, T^{\text{storage}}\}$, corresponding to the storage step, and the stochastic consumption time t^{consum}. Depending on the specific pathogen, Ξ_x comprises of other stochastic variables. For *Salmonella*, Ξ_x includes the rate parameters $\{\delta_{\text{core}}, \delta_{\text{rind}}\}$ and for *Lm* it includes the rate parameter ρ_{Lm}.

We define the batch risk $R_x^{\text{batch}}(\xi_x)$ as the probability of getting the particular illness from pathogen x, by consuming a portion of 25g of cheese from a

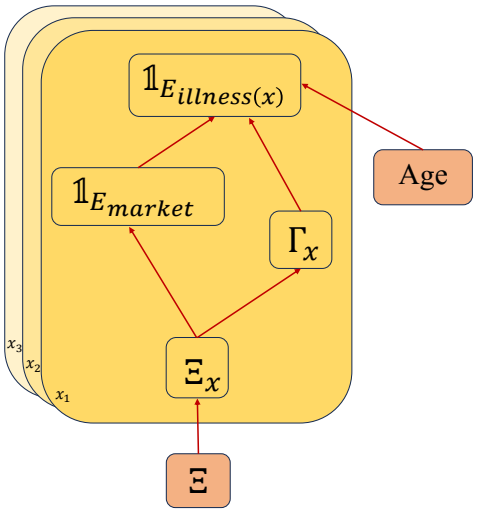


Figure 2.8: Directed acyclic graph showing the dependence between different model variables. The dependent variables are positioned at the arrow tips, signifying their dependence on the specific variables, which are situated at the arrow ends. No arrows between any two variables signifies independence. This graph is used to simplify the conditional expectations defined later in this chapter, using the dependence relationship.

particular batch, characterized by the internal variables $\Xi_x = \xi_x$. Further, we define two events $E_{\text{illness}(x)} \subset E_{\text{market}}$, where E_{market} indicates that the particular batch goes to the market (i.e., it is not rejected) and $E_{\text{illness}(x)}$ stands for the event that the particular batch goes to the market and the consumer gets the illness from pathogen x. The mutual dependence between different model variables and these events are shown in Figure 2.8. In this Directed Acyclic Graph, a specific dependent variable, conditioned by any of the variables placed at one of its arrow ends, is considered to be conditionally independent with respect to all the variables positioned at the subsequent arrow ends. From the definition of the batch risk, we have

$$\begin{aligned} R_x^{\text{batch}}(\xi_x) &= P\left[E_{\text{illness}(x)} \,|\, \Xi_x = \xi_x, E_{\text{market}}\right] \\ &= \mathbb{E}\left[\mathbb{1}_{E_{\text{illness}(x)}} \,|\, \Xi_x = \xi_x, \mathbb{1}_{E_{\text{market}}} = 1\right], \end{aligned} \tag{2.14}$$

and from the definition of the probability of illness in (2.11) - (2.13), we have

$$\begin{aligned} P_{\text{illness}(x)}(\text{age}, \gamma_x) &= P\left[E_{\text{illness}(x)} \,|\, \text{Age} = \text{age}, \Gamma_x = \gamma_x, E_{\text{market}}\right] \\ &= \mathbb{E}\left[\mathbb{1}_{E_{\text{illness}(x)}} \,|\, \text{Age} = \text{age}, \Gamma_x = \gamma_x, \mathbb{1}_{E_{\text{market}}} = 1\right]. \end{aligned} \tag{2.15}$$

Since $\mathbb{1}_{E_{\text{illness}(x)}}$ is conditionally independent of Ξ_x, given Γ_x, using (2.15) and

the law of total expectation, we can write

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{E_{\mathsf{illness}(x)}} \,|\, \Xi_x, \mathbb{1}_{E_{\text{market}}}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{E_{\mathsf{illness}(x)}} \,|\, \mathsf{Age}, \Gamma_x, \Xi_x, \mathbb{1}_{E_{\text{market}}}\right] \,|\, \Xi_x, \mathbb{1}_{E_{\text{market}}}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{E_{\mathsf{illness}(x)}} \,|\, \mathsf{Age}, \Gamma_x, \mathbb{1}_{E_{\text{market}}}\right] \,|\, \Xi_x, \mathbb{1}_{E_{\text{market}}}\right] \\
&= \mathbb{E}\left[P_{\mathsf{illness}(x)}(\mathsf{Age}, \Gamma_x) \cdot \mathbb{1}_{E_{\text{market}}} + 0 \cdot \mathbb{1}_{\overline{E}_{\text{market}}} \,|\, \Xi_x, \mathbb{1}_{E_{\text{market}}}\right].
\end{aligned} \tag{2.16}$$

Using (2.16) and the conditional independence with respect to E_{market} given Ξ_x, the definition of $R_x^{\text{batch}}(\xi_x)$ in (2.14) boils down to

$$\begin{aligned}
R_x^{\text{batch}}(\xi_x) &= \mathbb{E}\left[P_{\mathsf{illness}(x)}(\mathsf{Age}, \Gamma_x) \,|\, \Xi_x = \xi_x\right] \\
&= \sum_{\mathsf{age}=1}^{\mathsf{age}^{\max}} \int_0^\infty P_{\mathsf{illness}(x)}(\mathsf{age}, \gamma_x) \cdot p(\mathsf{age}, \gamma_x \mid \Xi_x = \xi_x)\, \mathsf{d}\gamma_x \\
&= \sum_{\mathsf{age}=1}^{\mathsf{age}^{\max}} g(\mathsf{age}) \int_0^\infty P_{\mathsf{illness}(x)}(\mathsf{age}, \gamma_x) p(\gamma_x|\xi_x) \mathsf{d}\gamma_x\,,
\end{aligned} \tag{2.17}$$

where the joint probability distribution of the Age and the dose $\Gamma_x \mid \Xi_x = \xi_x$ is $p(\mathsf{age}, \gamma_x \mid \Xi_x = \xi_x) = g(\mathsf{age}) \cdot p(\gamma_x|\xi_x)$, with $g(\mathsf{age}) = P[\mathsf{Age} = \mathsf{age} \mid E_{\text{market}}]$ being the age distribution of cheese consumers.

For MPS-STEC, the parameter $g(\mathsf{age})$ controls the proportion of cheese consumed by the age group (see, e.g., Perrin et al., 2014), and for *Salmonella*, the age parameter is considered to have no effects (Teunis, 2022, Teunis et al., 2010). The batch risk for *Lm* uses the model by Cadavez et al. (unpublished), where the population subgroups are based on different age groups for males and females: $\{[1,4],\ [5,14],\ [15,24],\ [25,44],\ [45,64],\ [65,74],\ [75,+\infty]\}$.

2.2.5 . Risk management options

2.2.5.1 Preharvest: milk sorting

The preharvest intervention strategy, a.k.a milk sorting, is carried out just before mixing the BTM from several farms into the ATM. In this step, the *E. coli* levels in BTM coming from the farms are tested and tanks with concentration above a certain threshold are rejected from the production chain. Each BTM is tested for *E. coli* concentration and the i-th farm is rejected if $Y_i^{\text{EC}} > l^{\text{sorting}}$. The milk loss for a particular batch is given by $M^{\text{batch}} = \sum_{i=1}^{N^{\text{farms}}} V_i \mathbb{1}_{\{i \notin S\}}$, where S denotes the set of farms that qualify after milk sorting. Note that the parameters controlling the preharvest intervention step are mainly the frequency of milk sorting f^{sorting} (measured in days) and the threshold of milk sorting l^{sorting} (measured in CFU/mL), as listed in Table 2.5. The milk sorting strategy is based on the hypothesis that *E. coli* and other pathogen strains follow the same fecal routes in the cows body (see, e.g. Perrin et al., 2014), as a

results the *E. coli* concentration in the BTM, can be used as a measure of the farms hygiene conditions.

Remark 2. In the current implementation of the QMRA model this milk sorting strategy only affects the concentration of STEC, MPS-STEC and *Salmonella* in the ATM. For *Lm* since the concentration in the ATM is simulated directly, it remains unaffected by the milk sorting. Implementation of a more realistic and efficient preharvest intervention strategy is discussed in Section 2.5.1.

2.2.5.2 Postharvest: microbial cheese sampling

The postharvest intervention strategy, a.k.a cheese sampling, can be implemented at different stages of cheese production depending on the type of bacteria. Typically, for raw milk soft cheese, the sampling process is carried out at the end of the salting step, during the third day of production. However, some producers may choose to implement cheese sampling during cheese ripening, until the 14-th day from the beginning of production. In our model, the cheese sampling time is determined by a parameter t^{test}, which is by default set at 14-th day of production. Cheese sampling consists in inspecting a batch of cheese for pathogen contamination, and this is done with respect to small portions, called sampling units, taken out of the batch. Once a single sample unit is tested positive for any of the three major pathogens under consideration, the whole batch of cheese is not sent to the market. Let n^{sample} and m^{sample} respectively denote the number of sample units taken from a batch and the mass of each sample unit (usually fixed at $25\,\text{g}$). The probability for a sample unit being tested positive for the three main pathogens, is

$$P_x^{\text{unit}}(\xi_x) = P\left(N_x^{\text{colony.sample}} > 0 \mid \Xi_x = \xi_x\right)$$

where conditionally on $\Xi_x = \xi_x$, $N_x^{\text{colony.sample}}$ is a Poisson distributed random variable, with expectation $(\lambda_x^{\text{colony}} \frac{m^{\text{sample}}}{w^{\text{cheese}}})$, and $\lambda_x^{\text{colony}}$ corresponds to the average number of colonies for pathogen x in the cheese sample, given $\Xi_x = \xi_x$.

For MPS-STEC and *Salmonella*, the average number of colonies $\lambda_x^{\text{colony}}$ is obtained as the sum of the average number of colonies of their respective subclass serotypes. The probability of detecting a colony in any of the sample units is $P_x^{\text{sample}}(\xi_x) = 1 - (1 - P_x^{\text{unit}}(\xi_x))^{n^{\text{sample}}}$. Assuming the detection events of the three types of pathogens are independent, the probability of rejecting the batch of cheese, given $\Xi = \xi$, where $\Xi = (\Xi_x)_x$, is obtained using the formula for the union of three independent events:

$$P^{\text{batch}}(\xi) = \sum_x P_x^{\text{sample}}(\xi_x) - \sum_{x<x^\dagger} P_x^{\text{sample}}(\xi_x) \cdot P_{x^\dagger}^{\text{sample}}(\xi_x) + \prod_x P_x^{\text{sample}}(\xi_x)\,. \tag{2.18}$$

The parameters for the postharvest step are listed in Table 2.9.

2.2.6 . Risk characterization

2.2.6.1 Output module

The output module, demonstrated in Figure 2.9, is outside the batch level simulator and its purpose is to obtain estimations of the expectation of the quantities of interest for a given set of $\theta = \{\theta^{\text{farm}}, \theta^{\text{cheese}}, \theta^{\text{con}}, \theta^{\text{post}}\}$. As will be explained in Section 2.3, the estimations will be carried out using Monte Carlo simulations. The output module uses three batch level simulator outputs, namely, the batch risk $R_x^{\text{batch}}(\xi_x)$, the probability of rejecting a batch $P^{\text{batch}}(\xi)$ and the milk loss M^{batch}.

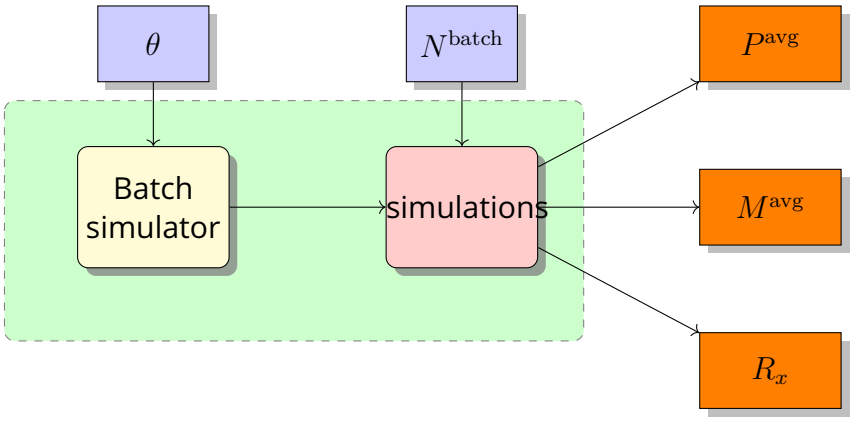


Figure 2.9: Output module denoted by the green box is the next hierarchical module of the batch level simulator. It simulates N^{batch} independent batches to estimate the average risk of illness R_x from pathogen x, average milk loss M^{avg} and average probability P^{avg} of rejecting a cheese batch.

The average batch rejection rate, or equivalently the average probability of a batch not going to the market, is defined as $P^{\text{avg}} = P[\overline{E}_{\text{market}}]$. Furthermore, we have $\mathbb{E}[\mathbb{1}_{\overline{E}_{\text{market}}} \,|\, \Xi = \xi] = p^{\text{test}} P^{\text{batch}}(\xi)$, where p^{test} is the proportion of cheese batches tested. Then using the law of total expectation, we can write

$$\begin{aligned} P^{\text{avg}} &= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\overline{E}_{\text{market}}} \,|\, \Xi\right]\right] \\ &= \mathbb{E}[p^{\text{test}} P^{\text{batch}}(\Xi)]\,. \end{aligned} \tag{2.19}$$

The average risk of illness R_x, from pathogen x, is the conditional probability $P[E_{\text{illness}(x)} \mid E_{\text{market}}]$, of getting the illness from consuming a portion of $25\,\text{g}$ of cheese from a batch of cheese, that was not rejected, produced with input parameters θ. Recalling $\Xi = (\Xi_x)_x$ and using the law of total expectation

twice, we can write

$$\begin{aligned} P[E_{\text{illness}(x)}] &= \mathbb{E}[\mathbb{1}_{E_{\text{illness}(x)}}] \\ &= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{E_{\text{illness}(x)}} \mid \Xi_x, \mathbb{1}_{E_{\text{market}}}\right]\right] \\ &= \mathbb{E}[R_x^{\text{batch}}(\Xi_x) \cdot \mathbb{1}_{E_{\text{market}}} + 0 \cdot \mathbb{1}_{\overline{E}_{\text{market}}}] \\ &= \mathbb{E}\left[R_x^{\text{batch}}(\Xi_x) \cdot \mathbb{E}\left[\mathbb{1}_{E_{\text{market}}} \mid \Xi\right]\right] \\ &= \mathbb{E}\left[R_x^{\text{batch}}(\Xi_x) \cdot (1 - p^{\text{test}} P^{\text{batch}}(\Xi))\right]. \end{aligned} \tag{2.20}$$

Using (2.20) and recalling $E_{\text{illness}(x)} \subset E_{\text{market}}$, we can write

$$\begin{aligned} R_x &= \frac{P[E_{\text{illness}(x)}]}{P[E_{\text{market}}]} \\ &= \frac{\mathbb{E}[R_x^{\text{batch}}(\Xi_x) \cdot (1 - p^{\text{test}} P^{\text{batch}}(\Xi))]}{1 - P^{\text{avg}}}. \end{aligned} \tag{2.21}$$

The average milk loss is defined as $M^{\text{avg}} = \mathbb{E}[M^{\text{batch}}]$.

Module outputs The three quantities of interest produces by the output module are, the average milk loss M^{avg}, average batch rejection rate P^{avg} and the average risk of illness R_x from pathogen x. These quantities of interest characterizes a cheese production process with input parameters θ.

2.2.6.2 DALY: Disability-Adjusted Life Years

A key feature of our multipathogen QMRA model is its ability to assess the impact of three pathogens on public health due to the consumption of raw milk soft cheese. The idea is to identify the most hazardous pathogen, and also to evaluate the collective impact of all pathogens on public health, specifically in terms of the years of healthy life lost due to the illnesses resulting from cheese consumption.

To assess the impact on public health from the potential illnesses induced by cheese consumption, we adopt an approach based on DALYs (Murray and Lopez, 1997), an acronym for Disability-Adjusted Life Years. DALYs combine the years of life lived with a disability (YLD) and the years of life lost (YLL) due to premature death from the illness.

The DALY takes into account a spectrum of symptoms, which include death, loss of kidney function, prolonged or short-term hospitalization, bloody diarrhea, diarrhea, among others. Furthermore, the age of the afflicted person impacts DALYs, with a child's death yielding higher DALYs compared to the death of an older individual.

DALY metric for cheese portions. First, we define the expected DALY due to illness(x) caused by the consumption of a cheese portion by a consumer of Age $=$ age,

$$\overline{\mathsf{DALY}}^{\text{portion}}_{\mathsf{illness}(x)}(\mathsf{age}) = \mathbb{E}[\mathsf{DALY}_x \mid \mathbb{1}_{E_{\mathsf{illness}_x}} = 1, \mathsf{Age} = \mathsf{age}], \tag{2.22}$$

where DALY_x is a random variable, denoting the DALY caused by the consumption of pathogen x. Next, we define the age specific average risk $R_x(\mathsf{age})$, from pathogen x, which is the conditional probability $P[E_{\mathsf{illness}(x)} \mid \mathsf{Age} = \mathsf{age}, \mathbb{1}_{E_{\text{market}}} = 1]$, for consumers with a particular age, of getting the illness from consuming a portion of 25 g of cheese from a batch of cheese, produced with input parameters θ and given that the particular batch actually goes into the market (not rejected). Following similar derivation steps of R_x, we can write

$$R_x(\mathsf{age}) = \frac{\mathbb{E}[R_x^{\text{batch}}(\Xi_x, \mathsf{age}) \cdot (1 - p^{\text{test}} P^{\text{batch}}(\Xi)) \mid \mathsf{Age} = \mathsf{age}]}{1 - P^{\text{avg}}}, \tag{2.23}$$

where the age specific batch risk $R_x^{\text{batch}}(\mathsf{age})$ is given by

$$R_x^{\text{batch}}(\xi_x, \mathsf{age}) = \int_0^\infty P_{\mathsf{illness}(x)}(\gamma_x, \mathsf{age}) \cdot p(\gamma_x \mid \Xi_x = \xi_x) \, \mathsf{d}\gamma_x. \tag{2.24}$$

Then, using the law of total expectation, the overall risk in (2.21) can be expressed as $R_x = \mathbb{E}[R_x(\mathsf{age})]$.

For the assessment of impact on public health, we are interested in the average DALY due to the consumption of cheese portions contaminated with pathogen x, from a batch of cheese that went to the market. This metric is denoted by $\overline{\mathsf{DALY}}_{\mathsf{portion},x}$, is derived using the law of total expectations:

$$\begin{aligned}
\overline{\mathsf{DALY}}_{\mathsf{portion},x} &= \mathbb{E}\left[\mathsf{DALY}_x \mid \mathbb{1}_{E_{\text{market}}} = 1\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathsf{DALY}_x \mid \mathsf{Age}, \mathbb{1}_{E_{\mathsf{illness}(x)}}, \mathbb{1}_{E_{\text{market}}}\right] \mid \mathbb{1}_{E_{\text{market}}} = 1\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathsf{DALY}_x \mid \mathsf{Age}, \mathbb{1}_{E_{\mathsf{illness}(x)}}\right] \mid \mathbb{1}_{E_{\text{market}}} = 1\right] \\
&= \mathbb{E}\left[\overline{\mathsf{DALY}}^{\text{portion}}_{\mathsf{illness}(x)}(\mathsf{Age}) \cdot \mathbb{1}_{E_{\mathsf{illness}(x)}} + 0 \cdot \mathbb{1}_{\overline{E}_{\mathsf{illness}(x)}} \mid \mathbb{1}_{E_{\text{market}}} = 1\right] \\
&= \mathbb{E}\left[\overline{\mathsf{DALY}}^{\text{portion}}_{\mathsf{illness}(x)}(\mathsf{Age}) \cdot \mathbb{1}_{E_{\mathsf{illness}(x)}} \mid \mathbb{1}_{E_{\text{market}}} = 1\right].
\end{aligned} \tag{2.25}$$

The expectation in (2.25) can be expressed as a weighted sum of age specific DALY values $\overline{\mathsf{DALY}}^{\text{portion}}_{\mathsf{illness}(x)}(\mathsf{age})$, with respect to $R_x(\mathsf{age})$ and $g(\mathsf{age})$,

$$\begin{aligned}
\overline{\mathsf{DALY}}_{\mathsf{portion},x} &= \sum_{\mathsf{age}=1}^{\mathsf{age}^{\max}} \overline{\mathsf{DALY}}^{\text{portion}}_{\mathsf{illness}(x)}(\mathsf{age}) \cdot \\
&\qquad P\left[E_{\mathsf{illness}(x)} \mid \mathsf{Age} = \mathsf{age}, \mathbb{1}_{E_{\text{market}}} = 1\right] \cdot P\left[\mathsf{Age} = \mathsf{age} \mid \mathbb{1}_{E_{\text{market}}} = 1\right] \\
&= \sum_{\mathsf{age}=1}^{\mathsf{age}^{\max}} \overline{\mathsf{DALY}}^{\text{portion}}_{\mathsf{illness}(x)}(\mathsf{age}) \cdot R_x(\mathsf{age}) \cdot g(\mathsf{age}).
\end{aligned} \tag{2.26}$$

The expectation in (2.25) can be expressed equivalently as

$$\begin{aligned}\overline{\mathsf{DALY}}_{\mathsf{portion},x} &= \sum_{\mathsf{age}=1}^{\mathsf{age}^{\max}} \overline{\mathsf{DALY}}^{\mathrm{portion}}_{\mathsf{illness}(x)}(\mathsf{age}) \cdot \\ &\qquad P\left[\mathsf{Age} = \mathsf{age} \mid E_{\mathsf{illness}(x)}\right] \cdot P\left[E_{\mathsf{illness}(x)} \mid E_{\mathrm{market}}\right] \\ &= \sum_{\mathsf{age}=1}^{\mathsf{age}^{\max}} \overline{\mathsf{DALY}}^{\mathrm{portion}}_{\mathsf{illness}(x)}(\mathsf{age}) \cdot \tilde{g}_\theta(\mathsf{age}) \cdot R_x \\ &= R_x \cdot \sum_{\mathsf{age}=1}^{\mathsf{age}^{\max}} \overline{\mathsf{DALY}}^{\mathrm{portion}}_{\mathsf{illness}(x)}(\mathsf{age}) \cdot \tilde{g}_\theta(\mathsf{age}),\end{aligned} \tag{2.27}$$

where $\tilde{g}_\theta(\mathsf{age}) = P\left[\mathsf{Age} = \mathsf{age} \mid E_{\mathsf{illness}(x)}\right]$ is the age distribution for the cases of illness caused by pathogen x, given the simulator inputs θ.

DALY estimation. Note that the estimation of the average DALY metric $\overline{\mathsf{DALY}}_{\mathsf{portion},x}$ using (2.26) or (2.27) includes the age specific DALY values denoted by $\overline{\mathsf{DALY}}^{\mathrm{portion}}_{\mathsf{illness}(x)}(\mathsf{age})$, which are not directly available in the literature. However, (2.27) can be simplified further under some assumptions.

The age distribution $\tilde{g}_\theta(\mathsf{age})$ for the illness(x) is dependent on the QMRA simulator inputs θ, through the corresponding dose-response model and is directly related to the proportion of cheese consumption $g(\mathsf{age})$ for different age group. Acknowledging that our dose-response models heavily rely on epidemiological data primarily sourced from food-borne illness outbreaks, we tentatively posit that the age distribution $\tilde{g}_\theta(\mathsf{age})$ specific to each illness remains relatively impervious to the influences of simulator inputs θ, for reasonable variations around their nominal (baseline) values. Under this assumption, the age distribution generated by our QMRA raw milk cheese simulator can be assumed to be closely aligned with the global pattern. Although we make the above assumption in this study, it is important to acknowledge that this hypothesis is made for simplification and modeling convenience. The validity of this hypothesis in the real-world context remains an open question and deserves further investigation.

For the calculation of the DALY metric, we reside to the study by Cassini et al. (2018), which is based on the Burden of Communicable Diseases in Europe (BCoDE) project (Kretzschmar et al., 2012, Mangen et al., 2013), focused on the EU/EEA population between 2009 and 2013. In this study, for each of the concerned illnesses, a model was created using the BCoDE toolkit (European Centre for Disease Prevention and Control, 2019). Within each model, age group-specific and sex-specific annual case numbers, multiplication factors to account for underestimation, and population data were input into the software. Finally a Monte Carlo simulation was performed with $10,000$ iterations for each illness, to estimate the median DALYs per case. Table 2.2 shows

the estimated median DALY corresponding to a singular case, represented as DALY(1 case)$_x$ along with the two components, namely, YLL and YLD for 1000 cases, for each associated illness. Given our assumptions, the average DALY metric in (2.27) can be simplified by substituting $\sum_{\text{age}=1}^{\text{age}^{\max}} \overline{\text{DALY}}^{\text{portion}}_{\text{illness}(x)}(\text{age}) \cdot \tilde{g}_\theta(\text{age})$ by DALY(1 case)$_x$ from Table 2.2,

$$\overline{\text{DALY}}_{\text{portion},x} = \text{DALY(1 case)}_x \times R_x. \tag{2.28}$$

Table 2.2: The estimated median values (Cassini et al., 2018) of years of life lived with disability (YLD) per 1000 cases, years of life lost (YLL) per 1000 cases, and DALY per 1 case, for STEC infections, listeriosis and salmonellosis. For MPS-STEC the DALY values are taken from ANSES (2020).

Pathogen	YLL (1000 cases)	YLD (1000 cases)	DALY(1 case)$_x$
MPS STEC infections	2700.0	1000.0	3.7
STEC infections	41.1	13.0	0.0541
Listeriosis	3300.0	400.0	3.7
Salmonellosis	15.0	4.0	0.019

It is important to highlight that, in our study, the values presented in Table 2.2 are used as approximations derived from the literature. This choice is motivated by the absence of pertinent analyses from epidemiological studies in the existing literature. Moreover, it serves the purpose of streamlining the computations within our study. However, it is crucial to recognize that utilizing global DALY values for assessing the health impacts of cheese consumption in France is not ideal. This approach relies on substantial assumptions, as elaborated earlier. Consequently, the computation of the DALY metric in our study does not fully capture the precise and prevailing impact on public health attributable to cheese consumption. Therefore, these results should be interpreted judiciously. Recognizing the necessity for greater accuracy, the computation of a more refined DALY metric incorporating relevant epidemiological studies is deferred as a prospective avenue for future research.

Moreover, our model considers the dose-response from Perrin et al. (2014), which estimates the risk of HUS, but the DALY values that could be found in the literature (see, e.g., Table 2.2) are either based on all symptoms of STEC or MPS-STEC. The investigation conducted by Lindqvist et al. (2023) identifies more predominant strains of MPS-STEC that specifically contribute to HUS, which is our primary focus. Therefore, a more comprehensive epidemiological study would be essential to estimate DALYs attributed solely to HUS. However, in this study we use the DALYs corresponding to MPS-STEC to compute the metric in (2.28).

The combined influence of the three pathogens, corresponding to the aggregated impact on public health from the consumption of raw milk cheese, is defined as $\overline{\text{DALY}}_{\text{portion}} = \sum_x \overline{\text{DALY}}_{\text{portion},x}$, ignoring the consequences of concurrent instances of two or more illnesses.

2.3 . Model implementation

2.3.1 . Mathematical and statistical techniques

2.3.1.1 Estimation of farm hygiene parameters

The hygiene level of each of the farms is characterized by two parameters, namely α and σ, which control the distribution of *E. coli* in the farm milk. In Perrin et al. (2014), the authors proposed a hierarchical Poisson mixed model to express the relationship between these parameters and the daily *E. coli* concentration x_d in BTM, that can be written as

$$\begin{aligned} x_d \mid \Lambda_d &\sim \text{Poisson}(\Lambda_d) \\ \log(\Lambda_d) &= \alpha + \varepsilon_d, \text{ where } \varepsilon_d \sim \mathcal{N}(0, \sigma^2)\,, \end{aligned} \tag{2.29}$$

where Λ_d denotes the average *E. coli* concentration in BTM of the farm on day d.

We propose a Bayesian approach based on Markov Chain Monte Carlo (MCMC) sampling to estimate the parameters of this model, separately for each farm. A detailed description of this approach can be found in Appendix A. *E. coli* test data was collected by CNIEL and ACTALIA from three French cheese producers, covering a total of 104 farms. The recorded data was lower truncated for observations less than 10 CFU/mL and upper truncated for observations above either 150 or 300 CFU/mL, depending on the producer.

Figure 2.10 plots two histograms of the values of α and σ^2 respectively, as estimated for all the farms. A positive correlation is observed in Figure 2.11, between the estimated values of the two hygiene parameters.

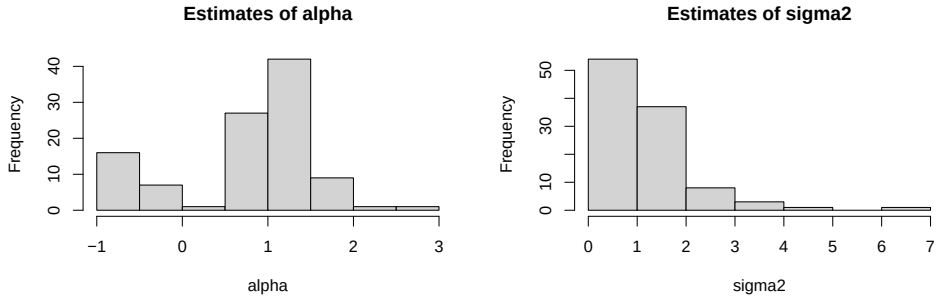


Figure 2.10: Estimated values of hygiene parameters, namely α and σ^2, using the data collected during $2019 - 2022$ from all the 97 farms (which had no missing data, out of 104 farms) under three different cheese producers in France.

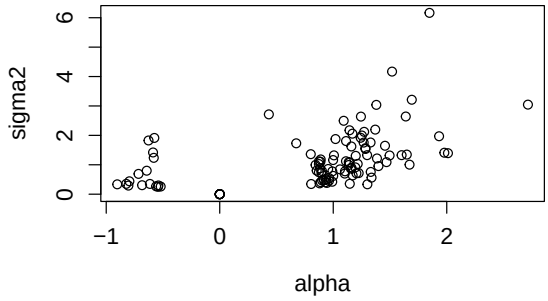


Figure 2.11: Scatter plot of the estimated values of α and σ^2

2.3.1.2 Computing $R^{\text{batch}}(\xi_x)$ using simple Monte Carlo

The computation of batch risk for *Lm* is addressed by the JEMRA (Cadavez et al., unpublished) toolbox in R, whereas for MPS-STEC and *Salmonella* one possible approach is based on the simple Monte Carlo integration method. It involves the computation of the integral in (2.17), with respect to the dose Γ_x conditional on the stochastic parameters ξ_x. This integral can be approximated using N^{dose} i.i.d samples $\{\gamma_{1,x}, \gamma_{2,x}, \ldots, \gamma_{N^{\text{dose}},x}\}$ from the conditional distribution $p(\Gamma_x \mid \Xi_x = \xi_x)$, of the random variable dose Γ_x, given the vector of stochastic internal variables ξ_x.

The approximated batch risk conditional on ξ_x is obtained as,

$$\hat{R}_x^{\text{batch}}(\xi_x) = \sum_{\text{age}=1}^{15} g(\text{age}) \frac{1}{N^{\text{dose}}} \sum_{i=1}^{N^{\text{dose}}} P_{\text{illness}(x)}(\text{age}, \gamma_{i,x}) \,, \tag{2.30}$$

where $P_{\text{illness}(x)}(\text{age}, \gamma_{i,x})$ is the probability of getting the illness from pathogen x, for the consumers of a particular age group, by consuming a portion of 25g of cheese with dose $\gamma_{i,x}$, coming from a batch associated with internal variables ξ_x.

2.3.1.3 Computing $R^{\text{batch}}(\xi_{\text{MPS-STEC}})$ using integral approximation

For MPS-STEC, the batch risk in (2.17) can be expressed as

$$R^{\text{batch}}_{\text{MPS-STEC}}(\xi_{\text{MPS-STEC}}) = \sum_{\text{age}=1}^{\text{age}^{\max}} g(\text{age})\, \mathbb{E}\left[(1-r_a)^{\Gamma_{\text{MPS-STEC}}}\right], \tag{2.31}$$

where $r_a = r_0 \cdot \exp(k \cdot \text{age})$. The random variable dose $\Gamma_{\text{MPS-STEC}}$ can be decomposed additively with respect to the two classes of MPS-STEC strains, $\Gamma_{\text{MPS-STEC}} = \sum_s \Gamma_s$ for $s \in \{\text{MPS O157:H7, MPS nonO157:H7}\}$, with

$$\Gamma_s = Y_s^{\text{consum}} N_s^{\text{colony}} 10^{\tau_{\epsilon_s}\epsilon} = d_s N_s^{\text{colony}} b_s^{\epsilon}, \tag{2.32}$$

where $d_s = Y_s^{\text{consum}}$, $b_s = 10^{\tau_{\epsilon_s}}$, and ϵ is a standard normal variable. The probability of getting HUS from each of the strains of MPS-STEC being independent, the expectation in (2.31) can be decomposed using the law of total expectations:

$$\begin{aligned}\mathbb{E}[(1-r_a)^{\Gamma_{\text{MPS-STEC}}}] &= \prod_s \mathbb{E}\big[\mathbb{E}[(1-r_a)^{\Gamma_s} \mid \epsilon]\big] \\ &= \prod_s \mathbb{E}\big[\mathbb{E}[(1-r_a)^{d_s N_s^{\text{colony}} b_s^{\epsilon}} \mid \epsilon]\big].\end{aligned} \tag{2.33}$$

The inner expectations $\mathbb{E}[(1-r_a)^{\Gamma_s} \mid \epsilon]$, can be analytically derived:

$$\begin{aligned}&\mathbb{E}[\,(1-r_a)^{N_s^{\text{colony}} d_s b_s^{\epsilon}} \mid \epsilon\,] \\ &= \sum_{n_s^{\text{colony}}=0}^{\infty} (1-r_a)^{n_s^{\text{colony}} d_s b_s^{\epsilon}} \exp(-\lambda_s^{\text{colony}}) \frac{(\lambda_s^{\text{colony}})^{n_s^{\text{colony}}}}{n_s^{\text{colony}}!} \\ &= \sum_{n_s^{\text{colony}}=0}^{\infty} \frac{\{(1-r_a)^{d_s b_s^{\epsilon}} \lambda_s^{\text{colony}}\}^{n_s^{\text{colony}}}}{n_s^{\text{colony}}!} \frac{\exp\left(-\{(1-r_a)^{d_s b_s^{\epsilon}} \lambda_s^{\text{colony}}\}\right)}{\exp\left(-\{(1-r_a)^{d_s b_s^{\epsilon}} \lambda_s^{\text{colony}}\}\right)} \exp(-\lambda_s^{\text{colony}}) \\ &= \exp(-\lambda_s^{\text{colony}}\{1-(1-r_a)^{d_s b_s^{\epsilon}}\}).\end{aligned} \tag{2.34}$$

Substituting (2.34) in (2.33), we get

$$\mathbb{E}[(1-r_a)^{\Gamma_{\text{MPS-STEC}}}] = \exp(-(\lambda_{\text{O157:H7}}^{\text{colony}} + \lambda_{\overline{\text{O157:H7}}}^{\text{colony}})) \prod_s \mathbb{E}[c_{s,1}^{c_{s,2}^{\epsilon_{s,3}}}], \tag{2.35}$$

where $c_{s,1} = \exp(\lambda_s^{\text{colony}})$, $c_{s,2} = (1-r_a)^{d_s}$ and $c_{s,3} = b_s$.

Given that $c_{s,1} > 1$, $c_{s,2} < 1$, and $c_{s,3} > 1$, the function $c_{s,1}^{c_{s,3}^{\epsilon}}$ is monotonically non-increasing with respect to ϵ. Such functions can be integrated using deterministic quadrature methods (e.g., the trapezoidal rule), which offer a better convergence rate compared to simple Monte Carlo (see, for example, Basak et al., 2022b).

2.3.1.4 Computation of simulator outputs

The final outputs of the simulator as mentioned in Section 2.2.6.1 are estimated using simple Monte Carlo method, by simulating N^{batch} independent batches.

$$\hat{M}^{\text{avg}} = \frac{1}{N^{\text{batch}}} \sum_{l=1}^{N^{\text{batch}}} M_l^{\text{batch}} \tag{2.36}$$

Each of the simulated batches yields $\{\xi_1, \xi_2, \ldots, \xi_{N^{\text{batch}}}\}$, the set of internal stochastic parameters which are used construct the unbiased estimated of the quantities of interest:

$$\hat{P}^{\text{avg}} = \frac{1}{N^{\text{batch}}} \sum_{l=1}^{N^{\text{batch}}} \hat{P}^{\text{batch}}(\xi_l) p^{\text{test}}. \tag{2.37}$$

$$\hat{R}_x = \frac{1}{N^{\text{batch}}(1 - \hat{P}^{\text{avg}})} \sum_{l=1}^{N^{\text{batch}}} \hat{R}_x^{\text{batch}}(\xi_l)(1 - \hat{P}^{\text{batch}}(\xi_l) p^{\text{test}}) \tag{2.38}$$

2.3.1.5 Computation of intervention cost

One of the several objectives of the ArtiSaneFood project includes the implementation of intervention strategies for controlling the risk of food-borne illness. In this case the cost of the intervention strategies plays an important role. As detailed in Section 2.2.5.1 and Section 2.2.5.2, the cheese production process involves two types of intervention strategies: one involving the testing of farm milk, and the other involving the testing of cheese after production. In addition to the expenses related to conducting the microbiological tests, these intervention steps involve the rejection of both farm milk and cheese batches. The QMRA simulator produces two outputs concerning the average loss of milk per batch M^{avg}, and the average probability of rejecting a batch of cheese P^{avg}. Thus, the total cost of the intervention steps (C) can be decomposed into two part corresponding to the preharvest ans postharvest steps, as $C = C^{\text{pre}} + C^{\text{post}}$:

$$\begin{aligned} C^{\text{pre}} &= N^{\text{farm}} C_{\text{test}}^{\text{milk}} (1/f^{\text{sorting}}) + M^{\text{avg}} C_{\text{loss}}^{\text{milk}} \\ C^{\text{post}} &= n^{\text{sample}} C_{\text{test}}^{\text{cheese}} p^{\text{test}} + P^{\text{avg}} p^{\text{test}} C_{\text{loss}}^{\text{cheese}} N^{\text{cheese}} \end{aligned} \tag{2.39}$$

The parameters $\{C_{\text{test}}^{\text{milk}}, C_{\text{loss}}^{\text{milk}}, C_{\text{test}}^{\text{cheese}}, C_{\text{loss}}^{\text{cheese}}\}$ corresponding to different intervention costs, are provided by French cheese producer representatives and N^{cheese} denotes the average number of cheeses present in a batch. All the parameters in (2.39) are described in Table 2.9.

2.3.2 . Programming tools

The multipathogen model is implemented in R and for STEC an implementation of the single pathogen version of the QMRA model is made available in Food Safety Knowledge Markup Language (FSK-ML) format to facilitate its reuse (Basak et al., 2024). This open format is based on predefined terms, metadata and controlled vocabulary to harmonize annotations of risk assessment models (see, e.g., Ungaretti Haberbeck et al., 2018). The FSKX implementation allows the user to execute the simulator on KNIME, using a set of input parameters listed in in Appendix B. By suitably adjusting the input parameter *cm_n_batch*, the user can run the FSKX implementation to either simulate a single batch (by setting *cm_n_batch* $= 1$) or multiple independent batches (by setting *cm_n_batch* > 1) to estimate the ultimate quantities of interest. Simulation of a single batch produces three numerical outputs, namely, the STEC concentration (CFU/ml) in milk put in production, the amount of milk loss (in liters) due to testing and the probability of rejecting the cheese batch. This also produces graphical representation of the evolution of STEC and colonies during cheese fabrication (both solid and liquid phase), and the evolution of the bacteria growth rate over the different phases of production. On the other hand, when multiple batches are simulated, it produces the estimates of the ultimate quantities of interest, averaged over these batches, namely, the relative risk of HUS computed with respect to a baseline scenario (with no intervention steps), the average milk loss (in liters) and the average probability of rejecting a cheese batch after production.The corresponding graphical outputs show the distribution of STEC concentration (CFU/ml) in ATM, the relative batch risk of HUS (computed with respect to the baseline risk), and the relative batch risk and batch rejection probability as a function of initial STEC concentration. The baseline scenario, i.e., the cheese production without any intervention step, can be simulated by appropriately choosing the parameters $p_{\text{test}} = 0$ and $f^{\text{sorting}} = \infty$.

2.3.2.1 Computational time

The current implementation of the multipathogen model takes around ~ 4.5 seconds, to simulate one batch. Table 2.3 lists the proportion of computational time (on a single core) spent while running the simulator. Evidently the cheese module takes more than 96% of the computational time, which is dedicated to the time required for solving the ordinary differential equations that models the growth of the pathogens, as shown in (2.8).

Table 2.3: Allocation of computational time (using a single core) over different modules of the QMRA simulator, corresponding to the simulation of single batch.

Module	Farm	Cheese	Consumer
Time spent	0.7%	96.4%	2.9%

The current implementation in R, uses the *ode45* function provided by the *pracma* (Borchers, 2022) package, that implements the Dormand-Prince $(4, 5)$ method. A detailed profiling on the computational time on the cheese module is shown in Table 2.4, which precises the proportion of time taken by each of the cheese production steps and their duration in the production process. The number of function evaluations performed by the ODE solver, depends on the duration of the cheese production step, the behavior of the ODE's solution within that duration, and the desired level of accuracy specified by the user. Table 2.4 shows the decomposition of the total time consumed by the ODE solver, while simulating a single batch. Except the second growth step of *Lm*, the other cheese production steps runs the ODE solver separately for three pathogens. The time taken is proportional to the total duration of the corresponding production step, however the second growth step spends relatively less time than others, despite of having a long duration. This can be explained by the behaviour of the function, or equivalently the maximum growth rate $\mu_{\mathrm{Lm}}^{\max}$ which is considered to be a constant (see, e.g., Section 2.5.1) for that particular step.

Table 2.4: Allocation of computational time among different steps of cheese production and their duration. The steps namely, storage, molding, draining and salting represents the total time required for all the three pathogens and shows the combined duration for the three pathogens as well. The second growth step only concerns *Lm* and shows the combined duration for the core and rind region.

Step	2nd growth (*L.mono*)	Storage	Molding	Draining	Salting
Time spent	34.9%	20.9%	4.9%	31.4%	7.9%
Duration (in h)	1136	36	9	51	13.5

2.3.3 . Data and parameters

This subsection lists all the parameters of the multipathogen model with their corresponding values are the references. Table 2.5 lists the parameters θ^{farm} corresponding to the farm module. Table 2.7 shows the frequency distribution of number of cows in different farms, which is used to simulate the number of cows in different farms. Table 2.6 lists the physico-chemical pa-

rameters required to compute the maximum growth rate $\mu^{\max}$ for the three different pathogens, and Table 2.10 lists the other cheese module parameters, which are collectively denoted as θ^{cheese}. Table 2.8 and Table 2.9 respectively lists the consumer module parameters θ^{con} and the postharvest module parameters θ^{post}.

Table 2.5: The inputs for the Farm module, collectively represented as (θ^{farm}), are categorized into two groups, visually separated by a dashed line. The first part includes inputs that the user can modify, while the second part comprises fixed parameters of the model.

Symbol	Description	Values
N^{farms}	Number of farms	31
N_i^{cow}	Number of cows in i-th farm	Table 2.7
q^{milk}	Avg. quantity of milk from a cow	25 Liters
f^{sorting}	Milk testing frequency	10 days
l^{sorting}	Max. limit of *E. coli* conc.	50 CFU/mL
α_i, σ_i	Hygiene parameter for i-th farm	Section 2.3.1.1
$a^{\text{weibull}}, b^{\text{weibull}}$	Param. of dist. of STEC in feces	$0.264, 16.288$ (Perrin et al., 2014)
$\mu^{\text{ecoli}}, \tau^{\text{ecoli}}$	Mean & sd of *E. coli* dist. in feces	$6, 0.3$ (Perrin et al., 2014)
μ_u, τ_u	Param. for estimating p_{STEC}	$-0.927, 1.47411$ (Perrin et al., 2014)
$\mu^{\text{Lm}}, \tau^{\text{Lm}}$	Parameters of *Lm* dist. in milk	$-7.178, 0.552$ (Sanaa et al., 2004)
$p_{\text{MPS-STEC}}$	Prop. of MPS-STEC carriers	0.025 (Perrin et al., 2014)
p_{Salmo}	Prop. of *Salmonella* infected cows	0.03 (Bonifait et al., 2021)
$p_{\text{HV-Salmo}}$	Prop. of HV *Salmonella* carriers	0.33 (Bonifait et al., 2021)

Table 2.6: Cardinal parameters for the computation of $\mu^{\max}$ for MPS-STEC (Perrin et al., 2014), *Salmonella* (Gonzales-Barron et al., 2022) and *Lm* (Sanaa et al., 2004).

Symbol	MPS-STEC	*Salmonella*	*Lm*
$T_{\min}$	5.5	3.4	-1.7
T_{opt}	40.6	38.5	37
$T_{\max}$	48.1	46	45.5
$\text{pH}_{\min}$	3.9	4	4.71
pH_{opt}	6.25	7	7.1
$\text{pH}_{\max}$	14	9	9.61
$\text{aw}_{\min}$	0.9533	0.94	0.913
aw_{opt}	0.999	0.99	0.997

Table 2.7: Distribution of number of cows: data provided by CNIEL and ACTALIA, collected from 31 producers of milk. This empirical distribution is used to simulate the number of cows in farm in the QMRA model.

Cows	$5-20$	$20-40$	$40-60$	$60-80$	$80-100$	$100-120$	Total
Farms	1	2	18	8	1	1	31

Table 2.8: Inputs of the consumer module collectively denoted as θ^{con}.

Symbol	Description	Values
k, r_0	Param. in STEC dose-response	$0.38, 1e{-}2.33$
$\text{age}^{\max}$	Maximum age group	15
w^{cheese}	mass of a single cheese	$250\,\text{g}$
w^{serving}	mass of a single serving	$25\,\text{g}$
$g(\text{age})$	Prop. of cheese consumed per age	(Perrin et al., 2014)
$\mu_{\epsilon_{\text{O157}}}$	Mean of ϵ_{O157}	0 (Basak et al., 2024)
$\tau_{\epsilon_{\text{O157}}}$	SD. of ϵ_{O157}	0.000279659 (Perrin et al., 2014)
$\mu_{\epsilon_{\overline{\text{O157}}}}$	Mean of $\epsilon_{\overline{\text{O157}}}$	0 (Basak et al., 2024)
$\tau_{\epsilon_{\overline{\text{O157}}}}$	SD. of $\epsilon_{\overline{\text{O157}}}$	0.000065399 (Perrin et al., 2014)
$\alpha_{\text{HV-Salmo}}$	HV *Salmonella* dose-response parameter	0.132 (Strickland et al., 2023)
$\alpha_{\text{LV-Salmo}}$	LV *Salmonella* dose-response parameter	0.318 (Strickland et al., 2023)
$\beta_{\text{HV-Salmo}}$	HV *Salmonella* dose-response parameter	51.45 (Strickland et al., 2023)
$\beta_{\text{LV-Salmo}}$	LV *Salmonella* dose-response parameter	4729.9 (Strickland et al., 2023)
$\theta_g,\ r$	*Lm* dose-response parameters	(Pouillot et al., 2015, Ricci et al., 2018)
N^{dose}	Monte Carlo sample size	10000

Table 2.9: Inputs for postharvest module collectively denoted as θ^{post} are shown in the first part of the table. In the second part the cost values of the intervention steps are shown.

Symbol	Description	Values
n^{sample}	Number of test portions	5
m^{sample}	Mass of each test portion	25 gm
p^{test}	Prop. of batch tested	0.5
t^{test}	Time when batch is tested	14th Day
$C^{\text{milk}}_{\text{test}}$	Cost of testing farm milk	10 EUR
$C^{\text{milk}}_{\text{loss}}$	Cost of rejecting one Liter milk	0.2 EUR
$C^{\text{cheese}}_{\text{test}}$	Cost of testing one cheese sample	70 EUR
$C^{\text{cheese}}_{\text{loss}}$	Cost of rejecting one cheese	1.5 EUR

Table 2.10: Cheese module inputs, collectively denoted as θ^{cheese}.

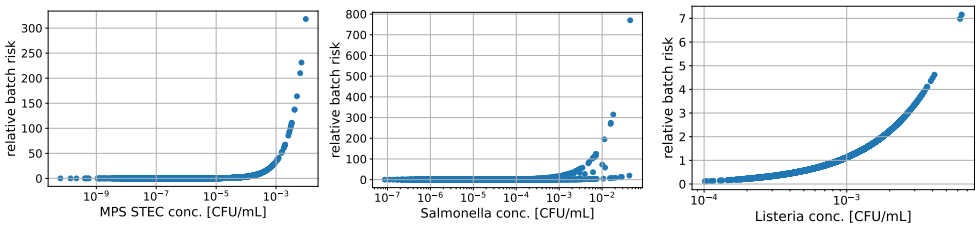
Symbol	Description	Values
$\theta(\mu^{\max})$	Cardinal params. for $\mu^{\max}$	Table 2.6
$\mu^{\text{opt}}_{\text{MPS-STEC}}$	Optimal growth rate for MPS-STEC	1.85 (Perrin et al., 2014)
$\mu^{\text{opt}}_{\text{Salmo}}$	Optimal growth rate for *Salmonella*	1.02 (Gonzales-Barron et al., 2022)
$\mu^{\text{opt}}_{\text{Lm}}$	Optimal growth rate for *Lm*	0.55 (Gonzales-Barron et al., 2022)
$y^{\max,\text{milk}}$	Hypothetical max population in milk	10^9 CFU/mL (Perrin et al., 2014)
$y^{\max,\text{cheese}}$	Hypothetical max population in cheese	10^5 CFU/g (Perrin et al., 2014)
p_{O157}	Class probability of MPS O157:H7	0.76 (Perrin et al., 2014)
$p_{\text{HV-Salmo}}$	Class probability of HV *Salmonella*	$Y^{\text{milk}}_{\text{HV-Salmo}}/(Y^{\text{milk}}_{\text{HV-Salmo}} + Y^{\text{milk}}_{\text{LV-Salmo}})$
$p_{\text{LV-Salmo}}$	Class probability of LV *Salmonella*	$1 - p_{\text{HV-Salmo}}$
$p^{\text{colony}}_{\text{HV-Salmo}}$	Consumption probability of HV *Salmonella*	$\lambda^{\text{colony}}_{\text{HV-Salmo}}/(\lambda^{\text{colony}}_{\text{HV-Salmo}} + \lambda^{\text{colony}}_{\text{LV-Salmo}})$
$p^{\text{colony}}_{\text{LV-Salmo}}$	Consumption probability of LV *Salmonella*	$1 - p^{\text{colony}}_{\text{HV-Salmo}}$
ρ_{O157}	MPS O157:H7 decline rate	0.14 ($\log_{10}$CFU/day) (Perrin et al., 2014)
$\rho_{\overline{\text{O157}}}$	MPS nonO157:H7 decline rate	0.033 ($\log_{10}$CFU/day) (Perrin et al., 2014)
δ_{core}	Core *Salmonella* decline rate	1.4 (Gonzales-Barron et al., 2022)
$\tau^{\delta}_{\text{core}}$	Inter cheese variability of *Salmonella*	0.021568496 (Gonzales-Barron et al., 2022)
δ_{rind}	Rind *Salmonella* decline rate	3.1 (Gonzales-Barron et al., 2022)
$\tau^{\delta}_{\text{rind}}$	Inter cheese variability of *Salmonella*	0.022373347 (Gonzales-Barron et al., 2022)
p_{core}	Param. for core *Salmonella* decline	0.274 (Gonzales-Barron et al., 2022)
p_{rind}	Param. for rind *Salmonella* decline	2.7 (Gonzales-Barron et al., 2022)
ρ_{Lm}	*L. mono* decline rate	Triangular$(0.5, 1, 2)$ (Sanaa et al., 2004)
pH_{Lm}	pH for *L. mono* second growth phase	6 (ArtiSaneFood challenge tests)
T_{Lm}	Temp. for *L. mono* second growth phase	12.45 (ArtiSaneFood challenge tests)
pH^{storage}	Industrial params. for storage step	Table III in Perrin et al. (2014)
d^{storage}	Storage duration (hours)	Triangular$(1, 12, 40)$ (Perrin et al., 2014)
T^{storage}	Storage temperature (°C)	Uniform(4,6) (Perrin et al., 2014)
$\{d, pH, T\}$	Physico-chemical parameters	Table III in Perrin et al. (2014)
aw	Parameter for water activity	0.99, Table III in Perrin et al. (2014)
v^{cheese}	Milk used in a single cheese	2200 mL (Basak et al., 2024)
w^{loss}	Proportion of water loss in molding	0.9 (Perrin et al., 2014)
t^{consum}	Consumption time	Triangular$(22, 30, 60)$ (Basak et al., 2024)

2.4 . Model evaluation

2.4.1 . Batch level outputs

As mentioned in Section 2.2.6.1, the batch level simulator produces three outputs, namely, the batch risk $R^{\text{batch}}_x(\xi_x)$, the probability of detecting contamination $P^{\text{sample}}_x(\xi_x)$ while testing and the milk loss per due to preharvest testing M^{batch}. These quantities are computed conditionally on the stochastic parameters ξ_x of the model. To compute the relative batch risk for each pathogen, the risk associated with the consumption of a particular batch of cheese, denoted by $R^{\text{batch}}_x(\xi_x)$, is divided by a baseline risk value R^{baseline}_x. R^{baseline}_x is the average risk of illness from pathogen x in a baseline scenario which corresponds to a specific situation with no preharvest or postharvest intervention steps. The relative batch risk against the initial concentration of pathogen in milk is plotted in Figure 2.12. 1000 independent batches are simulated with fixed values of the internal stochastic variables, with $d^{\text{storage}} = 12$ hours, $T^{\text{storage}} = 5$ degree Celsius, $t^{\text{consum}} = 30$ days, $\rho_{\text{Lm}} = 1$, $\delta_{\text{core}} = 1.4$ and $\delta_{\text{rind}} = 3.1$. This shows the monotonically increasing relationship with

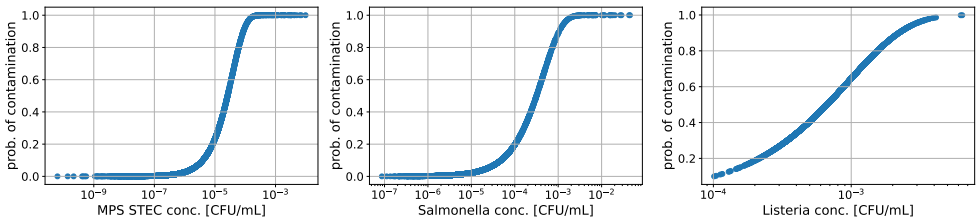
respect to Y_x^{milk}. For *Salmonella* the computation of risk still includes some extent of randomness due to the use of Monte Carlo method as explained in Section 2.3.1.2.



(a) STEC (b) *Salmonella* (c) *Lm*

Figure 2.12: Relative batch risk against initial concentration of pathogen in milk Y_x^{milk}, with fixed values of other internal variables in ξ_x.

In the same framework, the second output of interest the probability of detecting contamination $P_x^{\text{sample}}(\xi_x)$, i.e. detecting at least one colony at the time of testing n^{sample} sample units of mass m^{sample}g, is plotted in Figure 2.13, as a function of initial pathogen concentration in milk. Primarily Y_x^{milk} is the major influencing factor for determining the probability of detecting contamination as well. While MPS-STEC and *Salmonella* shows similar behaviour, due to high initial concentration, the probability of detection of contamination is higher for *Lm*.



(a) STEC (b) *Salmonella* (c) *Lm*

Figure 2.13: Probability of detecting contamination at the time of testing, against initial concentration of pathogen in milk Y_x^{milk}, with fixed values of other internal variables in ξ_x.

2.4.2 . Concentration and prevalence

The model was used to compute the prevalence of the three pathogens during different stages of cheese production. More precisely, the concentration of the pathogens in the ATM and the corresponding summary statistics over different batches were simulated, as listed in Table 2.11, in presence of no preharvest intervention. Based on the assumptions of the farm module

stated in Section 2.2.3.1, a baseline scenario with no milk testing step was simulated $10,000$ times to obtain these prevalence values.

Remark 3. It is to be noted that these figures does not represent the actual scenario of farm milk contamination in France, since these values are based on a hypothetical baseline scenario with no intervention steps, which is not the case in reality. The baseline scenario is used in this study as a reference to compare the impact of the intervention strategies.

Table 2.11: Baseline results of QRMA model: Summary statistics (mean, median, standard deviation and quantiles of order 5% and 95%) of the simulated concentration (in $\log_{10}$ scale) for STEC, MPS-STEC, *Salmonella* and *Lm*, in ATM with no milk testing intervention step.

Symbol	Mean	Median	SD	$q_{0.05}$	$q_{0.95}$
$Y^{\text{milk}}_{\text{STEC}}$	-3.44	-3.5	0.58	-4.28	-2.39
$Y^{\text{milk}}_{\text{MPS-STEC}}$	-5.91	-5.94	1.0	-7.51	-4.24
$Y^{\text{milk}}_{\text{HV-Salmo}}$	-5.78	-5.86	0.94	-7.21	-4.12
$Y^{\text{milk}}_{\text{LV-Salmo}}$	-5.2	-5.28	0.82	-6.41	-3.72
$Y^{\text{milk}}_{\text{Lm}}$	-3.12	-3.12	0.24	-3.5	-2.73

The prevalence of the pathogens in the cheese (in a standard serving of 25 g) was also simulated at the time of consumption. The prevalence is defined as the probability of observing at least one colony of the particular pathogen in the food-item. The prevalence is computed using the expected number of colonies given by (2.9), adjusted with respect to the size of the colonies at the time of consumption. In a baseline scenario $10,000$ batches were simulated with different consumption time, to compute the average prevalence in a cheese serving, as listed in Table 2.12.

Table 2.12: Baseline prevalence summary statistics (in percentage) computed at time of consumption, for cheese servings (25 g) with no intervention steps.

Symbol	Mean	Median	SD	$q_{0.05}$	$q_{0.95}$
MPS-STEC	1.97	0.16	8.21	0	7.43
HV *Salmonella*	0.37	0.02	3.22	0	0.91
LV *Salmonella*	0.75	0.06	4.67	0	2.27
Lm	39.47	37.03	21.97	8.37	80.03

Evidently the high prevalence of *Lm* can be attributed to two major reason, firstly the high concentration in ATM and as shown in Table 2.11. Secondly, as described in Section 2.2.3.2, *Lm* has a secondary growth phase during the ripening step of cheese production, which continues until the time of

consumption. The prevalence for all the three pathogens directly affects the postharvest sampling plan.

Consider a postharvest sampling plan with $n^{\text{sample}} = 5$ sample units of mass $m^{\text{sample}} = 25$ g, for each of the pathogens, such that the sample units are taken at the end of the ripening phase, i.e. at the 14-th day of production. Table 2.13 lists the expected probability $\mathbb{E}[P_x^{\text{sample}}(\Xi_x)]$ of detecting a colony of the respective pathogen, in any one of the sample units. This is estimated using the average probability of rejection over $10,000$ simulated batches with no preharvest intervention step.

Table 2.13: Expected probability $\mathbb{E}[P_x^{\text{sample}}(\Xi_x)]$ of detecting at least one colony in any of $n^{\text{sample}} = 5$ sample units of mass $m^{\text{sample}} = 25$ g.

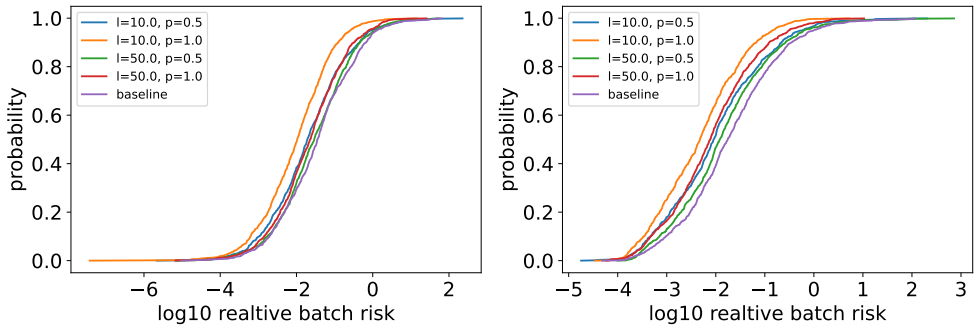
MPS-STEC	*Salmonella*	*Lm*
0.14	0.1	0.5

Clearly the high prevalence of *Lm* has a significant effect on the final output of the postharvest sampling plan, i.e. the probability of rejecting the batch $P^{\text{batch}}(\xi)$ for detecting at least one of the pathogen colonies, in any one of the sample units. A particular batch will have a high probability of getting rejecting due to the high prevalence of *Lm*. Implementation of a more realistic postharvest sampling plan addressing this issue is discussed in Section 2.5.1.

2.4.3 . Impact of intervention

As discussed in Section 2.2.5.1 and Section 2.2.5.2, there exist two types of intervention steps in the cheese production process, and we are interested in the finding the optimal values of the intervention parameters. The impact of the two types of intervention was studied qualitatively, using a series of different intervention scenarios defined using different combinations of intervention parameter values. More specifically, the effect of different intervention scenarios on the relative batch risk, for the preharvest step on and on the probability of batch rejection, for the postharvest step, were monitored.

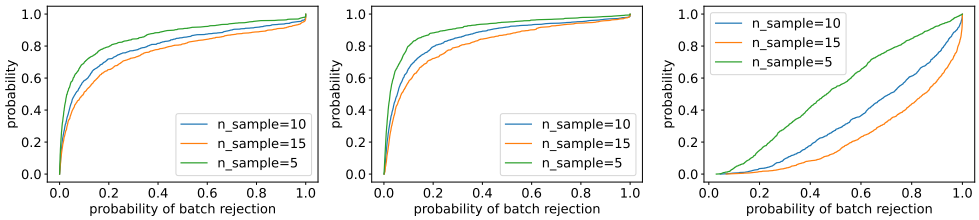
Figure 2.14 shows the impact of different scenarios of the preharvest intervention steps on the relative batch risk, only for two pathogens, MPS-STEC and *Salmonella*. Four simple scenarios were considered, implementing four different preharvest intervention strategies, with different values of the milk sorting limit $l^{\text{sorting}} = \{10, 50\}$, and different probability of testing a particular batch of milk $p = 1/f^{\text{sorting}} = \{0.5, 1\}$. Corresponding to each of the intervention scenario, 1000 independent batches were simulated to compute the empirical cumulative distribution functions (ECDF) of the relative batch risk. Evidently, the most strict scenario with a sorting limit at 10 CFU/ml and with a probability of testing each batch has the leftmost ECDF curve for the relative batch risk, as shown by the orange curve in Figure 2.14.



(a) MPS-STEC (b) *Salmonella*

Figure 2.14: Effect of preharvest intervention: empirical cumulative distribution functions of relative batch risk, corresponding to different intervention scenarios (varying milk sorting limit l and probability of batch testing p) and the baseline scenario. The leftmost curve corresponds to the most effective intervention strategy.

Remark 4. Here, we study the effect of preharvest intervention only on MPS-STEC and *Salmonella* since the proposed QMRA model uses directly the concentration of *Lm* in ATM, which remains unaffected by the preharvest intervention step (see Section 2.5.1).



(a) MPS-STEC (b) *Salmonella* (c) *Lm*

Figure 2.15: Effect of postharvest intervention: empirical cumulative distribution functions of probability of detecting a colony in any of the sample units, corresponding to different intervention scenarios with varying n^{sample}. The leftmost curve corresponds to the intervention strategy with highest probability of rejecting a cheese batch.

Figure 2.15 shows the impact of postharvest intervention step on the probability of detecting a colony in any of the sample units, if it is tested separately for each of the pathogen. The ECDFs are used to study qualitatively, the effect of number of cheese samples a.k.a sample units taken for the postharvest testing of cheese. The number of sample units considered, were $n^{\text{sample}} = \{5, 10, 15\}$, each of mass $m^{\text{sample}} = 25$g, drawn at the end of ripening step. For each of the postharvest scenarios, 1000 independent batches were simulated,

and the probability of detecting a colony in any of the sample units $P_x^{\text{sample}}(\xi_x)$ was computed separately for three different pathogens. Certainly, a higher number of sample units tend to produce chance of detecting contamination and, in turn, a higher probability of rejecting the batch. For MPS-STEC and *Salmonella*, the batch rejection probabilities were obtained in presence of no preharvest intervention, and for *Lm* it accounts for the particular preharvest intervention strategy as explained in Sanaa et al. (2004). Among the three pathogens, the ECDFs of *Lm* show higher chances of detecting contamination at testing, compared to the other two pathogens.

2.4.4 . Objectives of interest

The previous subsection studies qualitatively the effect of intervention parameters on the batch level outputs, namely, the relative batch risk and the probability of detecting contamination. However, the aim of this study is to find the optimal values of the intervention parameters that minimizes the two main objectives of the QMRA model, namely, the DALY per one portion $\overline{\text{DALY}}_{\text{portion}}$ and the total cost of intervention C. These two objectives are chosen in order to take into account the impact of intervention strategies both on the cheese consumers and the cheese producers. As demonstrated Section 2.4.3, strict intervention strategies can be helpful in reducing the risk of illness for the consumers, but at the same time it increases the probability of rejecting the milk and cheese batches. This trade off gives rise to a multiobjective optimization problem of a stochastic and computationally expensive simulator, which is formalized mathematically and addressed in Chapter 4 of this manuscript. To formalize the optimization problem, four intervention parameters corresponding to the preharvest and postharvest intervention steps, are considered. They are respectively, the frequency of testing the farm milk f^{sorting}, the threshold for milk testing l^{sorting}, the probability of testing a cheese batch p^{test} and the number of cheese sample units tested n^{sample}. The two objectives that are considered for the multiobjective optimization problem, are the the DALY per one portion of cheese $\overline{\text{DALY}}_{\text{portion}}$, caused by any of the pathogen and the total cost of intervention C, due to the loss of milk, cheese and analytical costs of testing. The computation of cost follows the computation of individual costs for different intervention steps, given by (2.39).

To visualize the trade-off between the two objectives $\overline{\text{DALY}}_{\text{portion}}$ and C, a series of different intervention scenarios are considered, using all possible combinations of values of the intervention parameters given in Table 2.14. A total of 216 different scenarios are constructed with these values and the QMRA model was evaluated for each of them with a Monte Carlo batch size 5000. Figure 2.16, demonstrates the relation between the two objectives, using scatter plots on the objective space. Colors are used to indicate the val-

Table 2.14: Values of intervention parameters to construct different intervention scenarios.

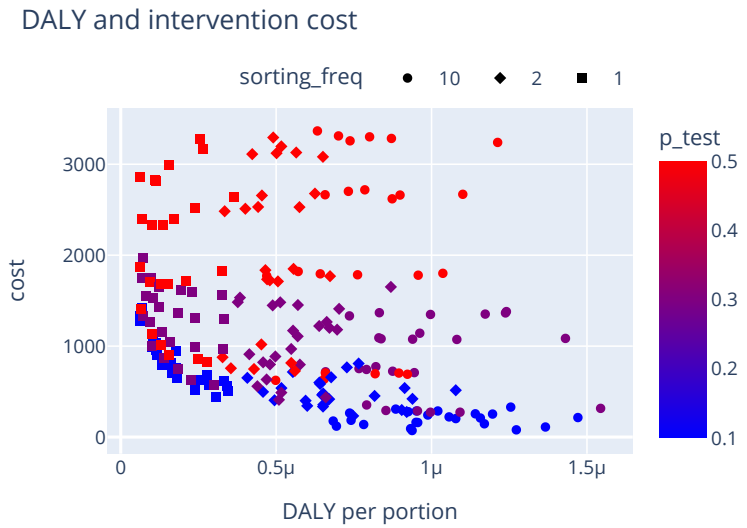
Parameter	Values	Units
f^{sorting}	$1,\ 2,\ 10$	Days
l^{sorting}	$10,\ 20,\ 30,\ 50,\ 100,\ 200$	CFU
p^{test}	$0.1,\ 0.3,\ 0.5$	Proportion
n^{sample}	$1,\ 5,\ 10,\ 15$	Sample units

ues of the cheese testing parameter p^{test} and different symbols are used to indicate the values of the milk testing parameter f^{sorting}. The left figure illustrates the scenario within a dual-pathogen framework, focusing solely on MPS-STEC and *Salmonella*, with both objectives considering only the impact of these two pathogens. The figure on the right incorporates the effects of all the three pathogens. The color transition from blue to red, signifying a stricter postharvest intervention, clearly shows a increase in the cost, however it is not so effective in reducing the the other objective DALY. This behaviour can attributed to the adapted strategy for rejecting a particular batch of cheese, in the postharvest sampling scheme, as described in Section 2.2.5.2. Unlike MPS-STEC, for the other two pathogens the rejection rule is not based on their corresponding high virulent or highly pathogenic strains. As a result due high prevalence, as shown in Table 2.12, pathogens like *Lm* are easily detectable when tested, but rather a smaller proportion of the entire population is actually pathogenic, which impacts the DALY. This explains why the effect of postharvest intervention on the DALY, is even less or almost null in the three-pathogen framework. On the other hand, the preharvest intervention parameter f^{sorting} as indicated by the symbols, shows an increasing impact on the cost and decreasing impact on the DALY metric. This explains the effectiveness of the preharvest intervention scheme on the two objectives of interest, though in the present model the concentration of *L. mono* remains unaffected by this intervention scheme.

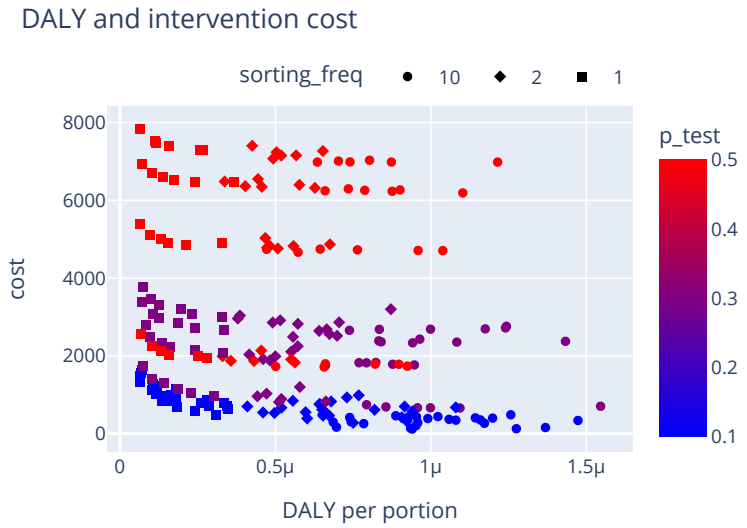
2.5 . Discussions and perspectives

2.5.1 . Model calibration

Model calibration or anchoring is a technique of improving the models efficiency, to be more compatible with observed data. The proposed model is intended to replicate the real life scenario of fabrication of a batch of cheese, however in reality not all practical gestures are possibly reproduced in the model. In this subsection, we list a few perspectives to be adapted in the current model, which are left as future work.



(a) MPS-STEC and *Salmonella*



(b) MPS-STEC, *Salmonella* and *Lm*

Figure 2.16: Figures showing the objective space consisting of the cost (in Euros) of intervention and the DALY per one portion (in $\mu := 1e-6$ scale), respectively for a two pathogen (MPS-STEC and *Salmonella*) and three pathogen (MPS-STEC, *Salmonella* and *Lm*) framework. The scatter plots were obtained with 216 different scenarios made up of different combinations of four intervention parameters $\{f^{\text{sorting}}, l^{\text{sorting}}, p^{\text{test}}, n^{\text{sample}}\}$, among which the parameters, f^{sorting} and p^{test} are indicated using different symbols and colors, respectively.

Preharvest intervention strategy The current model implements a simple preharvest intervention strategy based on the concentration of *E. coli* in the farms BTM. In other words the acceptance or rejection of farms is based on the *E. coli* concentrations which is assumed to be positively correlated with the concentrations of other pathogenic serotypes, as explained in Section 2.2.5.1. Although the farms BTM is tested for all the three types of pathogenic contamination, only the *E. coli* concentration can be measured quantitatively. Whereas for the other pathogens, due to low concentrations, only their presence can be detected and the strains can be identified. However in reality preharvest intervention strategies based on contamination of farm milk by *Salmonella* and *Lm* are practiced. To enhance the preharvest intervention strategy further, we can utilize this additional information, which is considered as a future perspective. Moreover, in the proposed QMRA model the concentration of *Lm* in the ATM, that is, the milk to be used for cheese production, is computed directly using 2.1. This was motivated by the work of Sanaa et al. (2004), who performed a Monte Carlo simulation study, based on the data collected in the years $2000 - 2001$, from respectively 347 and 79 farms, respectively, for Camembert and Brie in France. To compute the concentration of *Lm* in the milk used for cheese making, the authors followed the particular milk sorting strategy, for accepting the milk coming from different farms. The preharvest intervention step as explained in Section 2.2.5.1, is thus not applicable on the concentration of *Lm* and thereafter on the risk values, as simulated using the current version of our proposed model. The implementation of a more realistic and impactful preharvest step is left as a future perspective (see Section 2.5.1).

Computation of the DALY metric The computation of the DALY metric $\overline{\text{DALY}}_{\text{portion},x}$ as defined previously, is based on a series of assumptions which allows us to use the available DALY estimates from the literature (Cassini et al., 2018). This approach was adopted to simplify computations and it is to be noted that, the DALY metric thereby estimated is an approximate which requires careful interpretation and usage. A potential future perspective in this context includes using appropriate epidemiological studies to estimate the DALYs related to the consumption of raw milk cheese.

Pathogen classification based on virulence Pathogen classification based on virulence refers to categorizing microorganisms, into different groups or classes according to their ability to cause disease. Food-borne outbreaks tend to result from more virulent serotypes of the pathogens, which are commonly implicated in human illness. In the proposed model, high and low-virulence *Salmonella* (see, e.g., Strickland et al., 2023) serotype proportions were incorporated into each process module to account for their different im-

pact and exposure. For STEC the highly pathogenic and non pathogenic classification, was based on serotypes markers as proposed in Perrin et al. (2014). However, in Auvray et al. (2023) a more robust classification of pathogenic *E. coli* strains was proposed based on their potential virulence. The proposed model did not incorporate this new classification; however, it presents a promising avenue for enhancing the model in the future. For *Lm* there exists classification proportion of high, low and medium virulent serotypes for cold smoked salmon (see, e.g., Fritsch et al., 2018), but as far as our knowledge is concerned, these proportions remain unknown within the context of raw milk soft cheese in the existing literature. Developing a postharvest sampling plan centered around high-virulent serotypes could offer enhanced efficiency from an industrial perspective and provide a rational basis for decision-making.

Second growth step for Lm As shown in Figure 2.7, *Lm* has no decline phase after the salting step, and the colonies experience a second growth phase depending on the environmental conditions. The second growth phase is modeled using (2.8), which involves the maximum growth rate parameter $\mu_{\mathrm{Lm}}^{\max}(t)$, which depends on the physico-chemical parameters namely the pH, temperature and water activity. The dynamics of these environmental parameters are taken from Perrin et al. (2014), as shown in Figure 2.5a, and the available information extends only up to the salting step. According to challenge test data, it has been observed that the second growth step is activated when the pH increases to 6, from 4.52 starting from the end of salting step. Depending on the location of the colonies in the cheese matrix, the pH becomes favorable approximately after the 7-th and 20-th day of production, respectively for the rind and core region colonies. For simplicity, during the second growth phase, the proposed model uses a fixed value of the maximal growth rate $\mu_{\mathrm{Lm}}^{\max}(t)$, which is computed using pH 6, temperature 12.45 degree Celsius and water activity parameter 0.99. However a more realistic implementation can be achieved, by using a more dynamic approach to model the environmental physico-chemical parameters.

2.5.2 . Model validation and applicability

This work presents a QMRA model that offers a scientific approach to simulate the real-life scenarios encountered during the production of raw milk soft cheese. The model builds upon the previous work available in the literature, on the QMRA modeling of raw milk soft cheese (see, e.g. Basak et al., 2024, Perrin et al., 2014, Sanaa et al., 2004, Tenenhaus-Aziza et al., 2014) as well as expert opinions from ANSES, CNIEL, ACTALIA, and L2S. The primary goal of this type of model is to study the impacts of different process intervention parameters, in order to implement intervention strategies and make recommendations to cheese producers. However before deploying into in-

dustrial applications, careful precautions should be taken in terms of model quality assurance, that includes model verification, validation and calibration. Model verification includes checking the software code used to implement the model and providing a proper documentation. Future directions for this research work could involve the publication of the multipathogen model in the FSKX format, thereby promoting open access and facilitating easy and quick peer comparisons. The next crucial step is model validation which ensures the accuracy and reliability of the model's predictions. This involves assessing whether the QMRA model accurately reflects the real-world conditions and produces results that are consistent with observed data. The outputs of the proposed model are compared with previous QMRA works and published reports on contamination and outbreaks, and found to be consistent. For STEC the prevalence rates and baseline risk are compared to Perrin et al. (2014), for *Salmonella* the contamination rates were compared to reports published by the Fédération National des éleveurs de Chèvres (FNEC) in France, and for *Lm* the references from Food and Drug Administration et al. (2012), Sanaa et al. (2004) were compared. Despite of model quality assurance, the reliability and applicability of the model still remain subject to ongoing evaluation and refinement. According to World Health Organization et al. (2021) "Models are always incomplete representations of the system they are intended to model, but they can still be useful." Hence, it is essential to note that the outputs obtained using the simulator, such as the batch risk, loss of milk and proportion of rejecting cheese batches are just the estimates of a hypothetical scenario simulated with a state-of-the-art QMRA model. Depending on situations and model inputs these output can be significantly different from the actual prevalence of observed in reality. Nevertheless, the proposed multipathogen model continues to serve as a valuable tool for evaluating the efficacy of intervention strategies and aiding cheese producers in their decision-making processes.

3 - Integration of bounded monotone functions

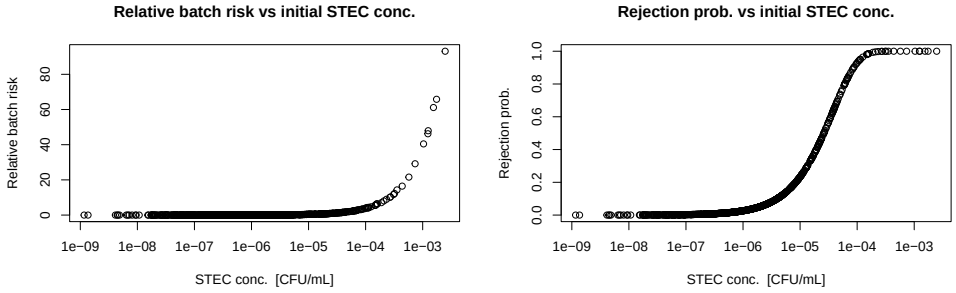
Section 3.2 of this chapter is a reproduction of the work by Basak et al. (2022b), with certain modifications.

3.1 . Introduction

3.1.1 . Problem and motivation

We address in this chapter the problem of constructing a numerical approximation of an integral $\mathbb{E}(g(Y)) = \int g(y)\,\mathrm{P}_Y(\mathrm{d}y)$, of a real, bounded and monotone function g. The integrating variable Y is a real-valued random variable, and the probability distribution P_Y is known. Using its cumulative distribution function F_Y, assuming its continuity, the domain of the integration problem can be transformed to $[0, 1]$ by applying the transformation $x = F_Y(y)$. Along with suitable scaling the problem boils down to the computation of $S(f) = \mathbb{E}(f(X)) = \int_0^1 f(x)\,\mathrm{d}x$, where X is uniformly distributed on $[0, 1]$ and f belongs to the class F of all non-decreasing functions defined on $[0, 1]$ and taking values in $[0, 1]$. All the results derived later in this chapter, assumes that the distribution of the integrating variable Y is known, and it is possible to reduce the integration problem in the above mentioned simplified form.

This study is primarily motivated by the application case of modeling the risk of food borne illness using a Quantitative Microbiological Risk Assessment (QMRA) model as described in Chapter 2. The QMRA model for raw milk soft cheese, can be considered as a stochastic simulator which produces outputs that are random variables. In our application case we work with a batch level simulator that produces the output corresponding to a batch of milk coming from a given set of farms, which is used for the fabrication of a single batch of cheese. The batch level simulator yields the batch risk R_x^{batch} for pathogen x and the batch rejection probability P^{batch}, as outputs. Figure 3.1 plots this two quantities for the pathogen MPS-STEC, but similar behaviour is observed for other pathogens as well. The two outputs are monotonically increasing with respect to the initial concentration of the pathogen in the farm milk. Moreover being a probability, both of these outputs are bounded as well. The ultimate quantities of interests of the QMRA model (e.g. the average risk of the illness and the average proportion of rejected cheese batches), are defined as functions of the expectations of R_x^{batch} and P^{batch}. A simple approach to compute these expectations is using simple Monte Carlo integration, which, however, turns out to be computationally expensive due to the batch size required to produce reliable estimates. We explore the possibility to construct possibly adaptive integration methods based on the monotonicity and boundedness



(a) Relative batch risk for MPS-STEC (b) Proportion of batch rejected

Figure 3.1: Figures depicting the variations in the outputs of the batch-level simulator as the initial concentration of MPS-STEC in the milk tank changes, while keeping other internal variables constant.

properties of the function, that is more efficient than the Monte Carlo approach. However, in the case of the QMRA application, the distribution of the integrating variable remains unknown, which is, in fact, quite common in many industrial applications. (see, e.g., Section 3.6). But, at the same time, it might be possible to simulate the integrating variable at a significantly lower cost than evaluating the function g itself, which motivates this study. In this chapter, we work in a fixed sample-size setting, where the number n of evaluations of f to be performed is chosen beforehand. The objective is to systematically study and summarize the existing methods within this particular framework and suggest alterations wherever applicable.

3.1.2 . Literature review

This problem was first studied by Kiefer (1957), who considered all the non-randomized methods, of both sequential and nonsequential types, where the sample points are chosen deterministically from the integration domain. kiefer showed that, in the set of all possible non-randomized methods based on n deterministic sample points, the one which achieves the minimum of the maximal numerical approximation errors, over the class F, turns out be a nonsequential method. A nonsequential integration method, also called a non-adaptive method, specifies in advance all values of the sample points at which the function is to be evaluated. This particular non-randomized (or, deterministic) method, mentioned by Kiefer (1957), takes regularly-spaced evaluations at $x_i = i/(n+1)$, $1 \leq i \leq n$, with $x_0 = 0$ and $x_{n+1} = 1$ and then uses the trapezoidal rule estimator, denoted by $\hat{S_n^{\mathrm{T}}}(f)$, which is essentially the average of the upper and lower Riemann sums of the intervals, constructed with respect to the regularly-spaced sample points.

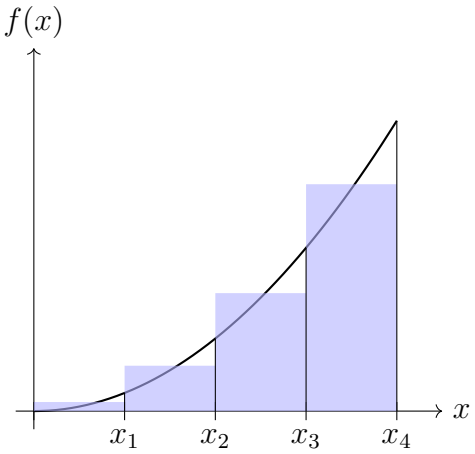


Figure 3.2: Trapezoidal rule estimate with average of Riemann sums.

$$\begin{aligned} \hat{S_n^{\mathrm{T}}}(f) &= \frac{1}{n+1} \sum_{k=1}^{n+1} \frac{1}{2}(f(x_k) + f(x_{k-1})) \\ &= \frac{1}{2n+2} + \frac{1}{n+1} \sum_{i=1}^{n} f(\frac{i}{n+1}), \end{aligned}$$

As notations, here we use the index $i = 1, 2, \ldots, n$, for the evaluation points and use the index $k = 1, 2, \ldots, (n+1)$, for the intervals. Inside the k-th interval the maximum approximation error due to the estimator $\hat{S_n^{\mathrm{T}}}(f)$ is exactly half of the area of difference between the two rectangles corresponding to the upper and lower Riemann sum, as shown in Figure 3.2. Let $p_k = 1/(n+1)$, denote the distance between the $(k-1)$-th and k-th sample point, for $1 \leq k \leq n+1$ and $\delta_k = f(x_k) - f(x_{k-1})$ be the difference in the corresponding functional values, then the maximum approximation error turns out to be $E = 1/2 \cdot \sum_{k=1}^{n+1} p_k \cdot \delta_k = (1/2) \cdot 1/(n+1) \sum_{k=1}^{n+1} \delta_k = 1/(2n+2)$. Novak (1992) later studied the problem with Monte Carlo (a.k.a. randomized) methods and showed that nonsequential Monte Carlo methods for the class F, have a lower bound on the maximal L_1 error of $1/8n$. As a consequence, Novak concluded that nonsequential randomized methods are not interesting for this special class F, however they can be useful for constructing sequential methods, that have much better convergence rate. Novak also studied sequential (a.k.a. adaptive) methods for which a particular sample point may depend on the previous sample values and the corresponding evaluations. He established that sequential methods are better in this setting than nonsequential ones, with a minimax rate of $n^{-3/2}$ over F for the L^1 error. Novak's proof relies on the construction of a particular two-stage algorithm, using stratified sampling

in the second stage.

3.1.3 . Outline

In this chapter we study both nonsequential and sequential randomized methods, with an objective to construct an efficient integration method for the class of function F. Section 3.2 is dedicated to the nonsequential randomized methods, which are key building blocks for the construction of good (rate-optimal) sequential methods—as can be learned from the proof of Theorem 3 in Novak's article. We derive a lower bound for the maximal L^p-error, for any $p \geq 1$, which is a generalization of a result by Novak (1992) concerning the L^1 error. Later in this section, we study the maximal L^2 error (variance) of two simple unbiased methods, based respectively on the control variate technique and on stratification. Section 3.3 is dedicated to sequential randomized methods, which are proved to be better in this setting than nonsequential ones, with a minimax rate of $n^{-3/2}$ over F for the L^1 error. We study the sequential method by Novak (1992) which is based on the construction of the nonsequential stratified sampling algorithm. We provide a tighter upper bound of the variance of Novak's estimator and propose certain modifications to the original algorithm, which in some special cases can be beneficial to reduce the unused sampling budget and provide better allocation of the sampling budget. In Section 3.4 we provide a set of numerical experiments over a set of different test problems, to assess the performance of the proposed modified version of Novak's algorithm. Section 3.5 demonstrates an application example on the Quantitative Microbiological Risk Assessment (QMRA) simulator. Finally Section 3.6 terminates the chapter with a discussion and the future work perspectives.

3.2 . Nonsequential randomized methods

A nonsequential Monte Carlo method first evaluates the function at n random points $X_1, \ldots, X_n$ in $[0, 1]$, and then approximates the integral $S(f)$ using an estimator

$$\hat{S}_n(f) = \varphi\left(X_1,\, f(X_1),\, \ldots,\, X_n,\, f(X_n)\right), \tag{3.1}$$

where $\varphi : [0, 1]^{2n} \to \mathbb{R}$ is a measurable function. A nonsequential method is thus defined by two ingredients: the distribution of $(X_1, \ldots, X_n)$ and the function φ. The class of nonsequential Monte Carlo methods as usually defined in the literature also allows $\hat{S}_n$ to be randomized (i.e., allows φ to be a random function). Since $\hat{S}_n$ is constructed using all the n sample points, which all together serves as a sufficient statistic for the integral estimation, from Rao-Blackwell's theorem it implies that considering randomized estimators in our definition is not interesting, for any convex loss function. In this context we

define the worst-case L^p error of such a method over the class F as,

$$e_p(\hat{S}_n) \;=\; \sup_{f\in F} \mathbb{E}\left(\left|S(f)-\hat{S}_n(f)\right|^p\right)^{1/p}. \tag{3.2}$$

3.2.1 . A lower bound for the maximal L^p error

Novak (1992) proved that for any nonsequential Monte Carlo method with sample size n, the maximal L^1 error $e_1(\hat{S}_n)$ is greater or equal to $1/(8n)$. We generalize this result to the case of the L^p error.

Lemma 3.2.1. *For any nonsequential Monte Carlo method with sample size n, and the integration domain divided into $2n$ equispaced intervals, there exists a particular interval I, containing no sample point with probability at least $1/2$.*

Proof of Lemma 3.2.1. Assume, for the sake of contradiction, for all the intervals the probability of being empty is < 0.5, or equivalently, for all the intervals the probability of being non empty is > 0.5. Then the expected number of sample points N_k in the k-th interval, for $k = 1, 2, \ldots, 2n$, is $\mathbb{E}[N_k] = \sum_{m=0}^{n} m \cdot P(N_k = m) \geq \sum_{m=1}^{n} P(N_k = m) > 1/2$. Now, the expected number of sample points in all the intervals is $\mathbb{E}[\sum_{k=1}^{2n} N_k] = \sum_{k=1}^{2n} \mathbb{E}[N_k] > 2n \cdot 1/2 = n$, which is a contradiction. □

Theorem 3.2.2. *For any nonsequential Monte Carlo methods with sample size n,*

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

Observe that Novak's lower bound is recovered for $p = 1$ and $p = 2$ gives the lower bound for the variance of unbiased nonsequential methods.

Corollary 3.2.3. *For any unbiased nonsequential Monte Carlo method with sample size n,*

$$\sup_{f\in F} \operatorname{var}\left(\hat{S}_n(f)\right) \;\geq\; \frac{1}{32n^2}.$$

Proof of Theorem 3.2.2. Consider a nonsequential Monte Carlo methods with evaluation points $X_1, \ldots, X_n$ and estimator $\hat{S}_n$. Divide the interval $[0, 1]$ into $2n$ equal subintervals of length $1/(2n)$: then using Lemma 3.2.1, there exists one subinterval, call it I, containing no evaluation point with probability at least $1/2$. Now construct two functions $f_1, f_2 \in F$ that are both equal to zero on the left of I, equal to one on the right, and such that $f_1 = 1$ and $f_2 = 0$ on I. Then $S(f_1) - S(f_2) = 1/(2n)$, and $\hat{S}_n(f_1) = \hat{S}_n(f_2)$ on the event $A = \{\{X_1, X_2, \ldots, X_n\} \cap I = \emptyset\}$, since f_1 and f_2 coincide outside of I. It follows that

$$\begin{aligned}\left(e_p(\hat{S}_n)\right)^p \;&\geq\; \sup_{f\in\{f_1,f_2\}} \mathbb{E}(|S(f)-\hat{S}_n(f)|^p) \;\geq\; \frac{1}{2}\sum_{j=1}^{2} \mathbb{E}(|S(f_j)-\hat{S}_n(f_j)|^p) \\ &\geq\; \frac{1}{2}\sum_{j=1}^{2} \mathbb{E}(|S(f_j)-\hat{S}_n(f_j)|^p \cdot \mathbb{1}_A) \;=\; \frac{1}{2}\sum_{j=1}^{2} \mathbb{E}(|S(f_j)-T|^p \cdot \mathbb{1}_A),\end{aligned}$$

where T denotes the common value of $\hat{S}_n(f_1)$ and $\hat{S}_n(f_2)$ on A. Now, using Jensen's inequality it can be proved that, for any $a, b, x \in \mathbb{R}$ and $p \geq 1$,

$$|a-x|^p + |b-x|^p \geq |a-b|^p / 2^{p-1}.$$

Then using this property we can write,

$$|S(f_1)-T|^p + |S(f_2)-T|^p \;\geq\; \frac{1}{2^{p-1}} \cdot |S(f_1)-S(f_2)|^p \;\geq\; \frac{1}{2^{p-1}} \cdot (\frac{1}{2n})^p$$

Thus we have,

$$\left(e_p(\hat{S}_n)\right)^p \;\geq\; \frac{1}{2} \cdot \frac{1}{2^{p-1}} \cdot (\frac{1}{2n})^p \cdot P(A) \;\geq\; \frac{1}{2^p} \cdot (\frac{1}{2n})^p \cdot \frac{1}{2} \;\geq\; (\frac{1}{2})^{2p+1} \cdot \frac{1}{n^p}.$$

□

3.2.2 . Uniform i.i.d. sampling

The simple Monte Carlo method is the most common example of a non-sequential method: the evaluation points $X_1, \ldots, X_n$ are drawn independently, uniformly in $[0,1]$, and then the integral is estimated by $\hat{S}_n^{\mathrm{MC}}(f) = \frac{1}{n}\sum_{i=1}^n f(X_i)$. The estimator is clearly unbiased, and it follows from the variance inequality by Popoviciu (1935)—i.e., $\mathrm{var}(Z) \leq 1/4$ for any random variable Z taking values in $[0,1]$—that

$$\left(e_2\left(\hat{S}_n^{\mathrm{MC}}\right)\right)^2 \;=\; \max_{f\in F} \mathrm{var}\left(\hat{S}_n^{\mathrm{MC}}(f)\right) \;=\; \frac{1}{4n}.$$

The maximal error is attained with the equality case in Popoviciu's inequality, i.e., when half of the probability is concentrated at the two bounds 0 and 1. This can be observed when f is a unit step function jumping at $x_0 = 1/2$, such that the evaluation point $f(X_i)$ has a Bernoulli distribution with probability $1/2$. It turns out that a smaller error can be achieved, for the same (uniform i.i.d.) sampling scheme, using the control variate technique. More specifically, we consider the control variate $\tilde{f}(X_i) = X_i$ and set

$$\hat{S}_n^{\mathrm{cv}}(f) \;=\; \frac{1}{n}\sum_{i=1}^n \left(f(X_i) - \tilde{f}(X_i)\right) + \frac{1}{2}.$$

Theorem 3.2.4. *The estimator $\hat{S}_n^{\mathrm{cv}}(f)$ is unbiased, and satisfies*

$$\left(e_2\left(\hat{S}_n^{\mathrm{cv}}\right)\right)^2 \;=\; \max_{f\in F} \mathrm{var}\left(\hat{S}_n^{\mathrm{cv}}(f)\right) \;=\; \frac{1}{12n}.$$

The maximal error is attained for any unit step function.

Proof. The estimator is unbiased since $\mathbb{E}(\tilde{f}(X_i)) = 1/2$, and therefore the mean-squared error is equal to $\mathrm{var}(\hat{S}_n^{\mathrm{cv}}(f)) = \frac{1}{n}\mathrm{var}(f(X)-X)$. For a unit step function $f = \mathbb{1}_{[x_0,1]}$ with a jump at $x_0 \in [0,1]$, the random variable $f(X)-X$ is

uniformly distributed over $[-x_0, 1-x_0]$, which yields $\mathrm{var}(\hat{S}_n^{\mathrm{cv}}(f)) = 1/(12n)$ as claimed. It remains to show that $\mathrm{var}(f(X)-X) \leq \frac{1}{12}$ for all $f \in F$.

Let $F_m \subset F$ denote the class of all non-decreasing staircase functions of the form $f = \sum_{k=1}^{m} \alpha_k \cdot \mathbb{1}_{(\frac{k-1}{m}, \frac{k}{m}]}$, with $0 \leq \alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_m \leq 1$. For any $f \in F$, consider the piecewise-constant approximation $f_m \in F_m$ defined by averaging f over each subinterval of length $1/m$, such that, $\mathbb{E}(f_m(X)) = \mathbb{E}(f(X))$. Then using Jensen's inequality it can be derived, for some suitable constant c, $|\mathrm{var}\left(f_m(X)-X\right) - \mathrm{var}\left(f(X)-X\right)| \leq c \cdot \mathbb{E}(|f_m(X)-f(X)|)$. By construction this quantity is upper bounded by the maximum approximation error of the trapezoidal rule $\mathbb{E}(|f_m(X)-f(X)|) \leq 1/2(m+1)$. Thus,

$$\sup_{f\in F} \mathrm{var}\left(f(X)-X\right) \;=\; \lim_{m\to\infty} \sup_{f\in F_m} \mathrm{var}\left(f(X)-X\right). \tag{3.3}$$

Let us now show that $\mathrm{var}\left(f(X)-X\right)$ is maximized over F_m when f is a unit step function. Pick any $f = \sum_{k=1}^{m} \alpha_k \cdot \mathbb{1}_{(\frac{k-1}{m}, \frac{k}{m}]} \in F_m$. Set $\alpha_0 = 0$ and $\alpha_{m+1} = 1$. If f is not a unit step function, then there exist $k_1, k_2 \in \{1, \ldots, m\}$ such that $k_1 \leq k_2$ and $\alpha_{k_1-1} < \alpha_{k_1} = \ldots = \alpha_{k_2} < \alpha_{k_2+1}$. Denote by $f_u \in F_m$ the function obtained by changing the common value of $\alpha_{k_1}, \ldots, \alpha_{k_2}$ in f to $u \in [\alpha_{k_1-1}, \alpha_{k_2+1}]$. The variance $\mathrm{var}(f_u(X)-X)$ is a convex function of u, since it can be expanded as au^2+bu+c with $a = \frac{k_2-k_1+1}{m}(1-\frac{k_2-k_1+1}{m}) > 0$. Consequently, we have $\mathrm{var}(f_u(X)-X) > \mathrm{var}(f(X)-X)$ at one of the two endpoints of $[\alpha_{k_1-1}, \alpha_{k_2+1}]$. Note that the corresponding staircase function f_u has one step fewer than f. Iterating as many times necessary, we conclude that for any $f \in F_m$ there exists a unit step function $g \in F_m$ such that $\mathrm{var}\left(f(X)-X\right) \leq \mathrm{var}\left(g(X)-X\right) = \frac{1}{12}$. Therefore $\sup_{f\in F_m} \mathrm{var}\left(f(X)-X\right) = \frac{1}{12}$, which completes the proof. □

3.2.3 . Stratified sampling

The use of stratification is widely popular in the literature of sample survey (see, e.g., Cochran, 1977). It was then adapted for use in Monte-Carlo estimation methods as a variance reduction technique (see, e.g., Hammersley and Handscomb, 1964, Kahn, 1956). Stratified sampling consists of partitioning the input space of X into mutually exclusive and exhaustive blocks called strata and then sampling independently from each stratum. Consider now a stratified sampling estimator with K strata:

$$\hat{S}_n^{\mathrm{str}}(f) \;=\; \sum_{k=1}^{K} w_k \cdot \frac{1}{n_k} \sum_{i=1}^{n_k} f\left(X_{k,i}\right), \tag{3.4}$$

where the k-th stratum is $I_k = [x_{k-1}, x_k]$, $0 = x_0 < x_1 < \cdots < x_{K-1} < x_K = 1$, the weight $w_k = |x_{k-1} - x_k|$ is the length of the k-th stratum, the allocation scheme $(n_1, \ldots, n_K)$ is such that $n_k \geq 0$ for all k and $\sum_k n_k = n$, and the random variables $X_{k,i}$ are independent, with the $X_{k,i}$s uniformly distributed in I_k.

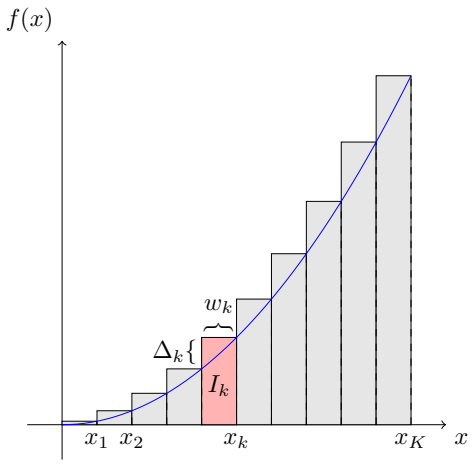


Figure 3.3: Stratified sampling with equispaced strata

Figure 3.3 demonstrates equispaced strata for a monotonically increasing function $f(x)$, with a particular stratum I_k, highlighted in red, along with the corresponding strata weight w_k and strata bound difference $\Delta_k = f(x_k) - f(x_{k-1})$. Note that the sampling points are no longer identically distributed here. However the samples inside each strata are i.i.d such that $\mathrm{var}(f(X_{k,i})) = \tau_k^2$, for all $i = 1, 2, \ldots, n_k$. The estimator $\hat{S}_n^{\mathrm{str}}(f)$ is unbiased, with variance

$$\mathrm{var}\left(\hat{S}_n^{\mathrm{str}}(f)\right) = \sum_{k=1}^{K} \frac{w_k^2 \cdot \tau_k^2}{n_k}. \tag{3.5}$$

Theorem 3.2.5. *For any $K \leq n$, any choice of strata and any deterministic allocation scheme, the stratified sampling estimator* (3.4) *satisfies*

$$\left(e_2\left(\hat{S}_n^{\mathrm{str}}\right)\right)^2 = \max_{f \in F} \mathrm{var}\left(\hat{S}_n^{\mathrm{str}}(f)\right) = \frac{1}{4} \max_k \frac{w_k^2}{n_k}. \tag{3.6}$$

The maximal error is attained for a unit step function with a jump at the middle of I_{k^*}, where $k^* \in \mathrm{argmax}\, w_k^2/n_k$. The minimal value of the maximal error (3.6) *is $\frac{1}{4n^2}$, and is obtained with $K = n$ strata of equal lengths ($w_k = n^{-1}$ and $n_k = 1$ for all k).*

The optimal stratified sampling method can be seen as a one-dimensional special case of the Latin Hypercube Sampling (LHS) method (McKay et al., 1979). (On a related note, McKay et al. (1979) prove that, in any dimension, the LHS method is preferable to the simple Monte Carlo method if the function is monotone in each of its arguments.)

Proof. For all $K \in \mathbb{N}^*$, let $\mathbf{\Delta}_K = \left\{(\Delta_1, \ldots, \Delta_K) \in \mathbb{R}_+^K \mid \sum_{k=1}^K \Delta_k \leq 1\right\}$. For a given stratified sampling method with K strata, for all $\underline{\Delta} \in \mathbf{\Delta}_K$, define

$$F_{\underline{\Delta}} = \left\{f \in F \mid \forall k \in \{1, \ldots, K\},\ f(x_k) - f(x_{k-1}) = \Delta_k\right\}.$$

Then it follows from (3.5) and Popoviciu's inequality that

$$\max_{f \in F_{\underline{\Delta}}} \operatorname{var}\left(\hat{S}_n^{\mathrm{str}}(f)\right) = \frac{1}{4} \sum_{k=1}^{K} \frac{w_k^2 \Delta_k^2}{n_k}, \tag{3.7}$$

where the maximum is attained for a non-decreasing staircase function with jumps of height Δ_k at the middle of the strata. Note that $\sum_{k=1}^{K} \Delta_k^2 \leq \sum_{k=1}^{K} \Delta_k \leq 1$. Therefore, the right-hand side of (3.7) is upper-bounded by $\frac{1}{4} \max_k w_k^2/n_k$, which is indeed the value of the variance (3.5) when f is a unit step function with a jump at the middle of the stratum where w_k^2/n_k is the largest.

In order to prove the second part of the claim, observe that any stratum with $n_k \geq 2$ can be further divided into n_k sub-strata of equal lengths without increasing the upper bound. Considering then the case where $K = n$ and $n_k = 1$ for all k, the upper bound reduces to $\frac{1}{4} \max_k w_k^2$, which is minimal when $w_1 = \cdots = w_n = n^{-1}$ since $\sum_k w_k = 1$. □

3.2.4 . Discussion

The stratified sampling (LHS) method provides the best-known variance upper bound over the class F for an unbiased nonsequential method as soon as $n \geq 3$, but is outperformed by the control variate method of when $n \leq 2$. We do not know at the moment if these results are optimal in the class of unbiased nonsequential methods. (The ratio between the best variance upper bound and the lower bound of Corollary 3.2.3 is $\frac{8}{3} \approx 2.67$ for $n = 1$, $\frac{16}{3} \approx 5.33$ for $n = 2$ and 8 for $n \geq 3$.)

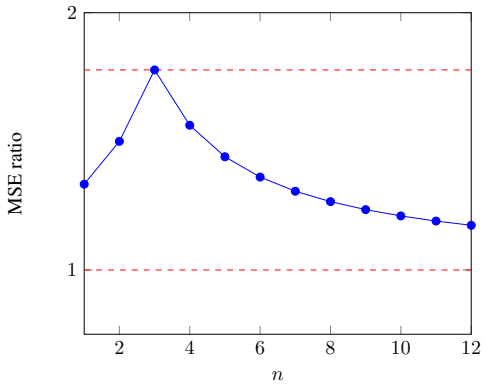


Figure 3.4: Ratio between the worst case MSE of the best known unbiased Monte Carlo method and the worst case squared error of the trapezoidal rule, for varying sampling budget. For $n = 1, 2$, the ratio is computed with respect to the control variate method, and starting from $n = 3$, the stratified sampling method is considered. Two horizontal red dotted lines are shown corresponding to the maximum (at ≈ 1.78) and limit value (at 1) of the ratio.

Relaxing the unbiasedness requirement, it turns out that both methods are outperformed for all n by the (deterministic) trapezoidal method discussed

in the introduction, which has a worst-case squared error of $1/(4(n+1)^2)$. The ratio of worst-case mean-squared errors, however, is never very large—at most $\frac{16}{9} \approx 1.78$, for $n = 3$—and decreases to 1 when n goes to infinity, as shown in figure 3.4.

As identified by Novak (1992), the nonsequential stratified sampling method can be used to construct a two-stage sequential randomized method with better convergence rate. The next section studies Novak's sequential algorithm and proposes several improvements.

3.3 . Sequential randomized methods

3.3.1 . Budget allocation for stratified sampling

Before studying sequential randomized methods, we first recall the budget allocation strategy for the stratified sampling method, described in the previous section. For a given function and given strata, the sampling budget allocation rule can be derived by minimizing the variance of the stratified sampling estimator $\mathrm{var}(\hat{S}_n^{\mathrm{str}}(f))$, subject to the total budget constraint $\sum_{k=1}^{K} n_k = n$. This constrained optimization problem is addressed using Lagrange's multiplier method (see, e.g., Cochran, 1977), that yields the sample sizes for each strata

$$n_k^{\mathrm{Neyman}} = \frac{w_k \tau_k}{\sum_{k=1}^{K} w_k \tau_k} \cdot n \tag{3.8}$$

This result (3.8) is also known as Neyman's allocation as it was obtained by Neyman (1934) in the literature of sample survey. It turns out that Tschuprow (1923) had obtained the result over a decade earlier, showing that the result follows as a special case of a more general problem. Substituting the values of n_k^{Neyman} in (3.5), we get the variance of the stratified sampling estimator with Neyman's allocation

$$\mathrm{var}(\hat{S}^{\mathrm{str}}_{n,\mathrm{Neyman}}(f)) = \frac{1}{n}\Big(\sum_{k=1}^{K} w_k \tau_k\Big)^2 \tag{3.9}$$

One of the problems with the allocation rule in (3.8) is that it can produce non-integer solutions and rounding off to nearest integer can either exceed the sampling budget, or leave a portion of the budget unutilized. In fact this problem is addressed in the literature of sample survey by Wright (2012, 2017). Another problem is, the allocation rule requires the strata variances τ_k^2 which are unknown. A classical solution is to perform pilot surveys to estimate these strata variances, and use them for deciding the allocation budget. This can be thought to be equivalent to constructing a two-stage sequential method (see, e.g., Novak, 1992), where the first step dedicated to to decide the allocation of the sampling budget. Later we will see that the allocation rule proposed by Novak (1992) is similar to the Neyman's allocation rule as in (3.8). Both of the

approaches are motivated by the idea of allocating more sampling budget to strata with higher variances.

3.3.2 . Novak's sequential algorithm

3.3.2.1 Algorithm description

In his article Novak (1992) proposed a two-stage sequential algorithm based on stratified sampling that achieves the minimax rate $O(n^{-3/2})$. Given a total sampling budget n, in the first step $m+1$ equispaced strata are constructed, by appropriately choosing $m = (n-1)/3$ equispaced sample points $x_i = i/(m+1)$, for $i = 1, 2, \ldots, m$, with $x_0 = 0$ and $x_{m+1} = 1$. The next step consists in allocating a sampling budget of n_k to each of the k strata, denoted by $I_k = [x_{k-1}, x_k]$, for $k = 1, 2, \ldots, (m+1)$, using the rule,

$$n_k = \begin{cases} 1 & \text{if } \Delta_k = 0 \\ \lceil (m+1) \cdot \Delta_k \rceil & \text{if } \Delta_k > 0, \end{cases} \tag{3.10}$$

where $\Delta_k = f(x_k) - f(x_{k-1})$ and $n_k = 1$ for $\Delta_k = 0$. Now for each strata instead of drawing n_k i.i.d samples from it, Novak uses stratification and further splits the k-th stratum into n_k equispaced substrata and then samples one sample from each of them. As studied in the previous section, the use of stratification in the second step gives a much lower worst case L^2 error of $1/(4 \cdot n_k^2)$ compared to $1/(4 \cdot n_k)$ in case of uniform i.i.d sampling with no stratification.

The construction of the unbiased estimator proposed as by Novak is similar to a stratified sampling estimator (see, (3.4)),

$$\hat{S}_n^{\text{novak}}(f) = \sum_{k=1}^{m+1} w_k \cdot \frac{1}{n_k} \sum_{i=1}^{n_k} f(X_{k,i}) \tag{3.11}$$

where, $X_{k,i}$ denotes the sample drawn from the i-th substratum inside I_k and $w_k = |x_k - x_{k-1}|$.

3.3.2.2 Justification of the budget allocation rule

The variance of Novak's estimator is obtained as,

$$\text{var}(\hat{S}_n^{\text{novak}}(f)) = \sum_{k=1}^{m+1} \frac{w_k^2}{n_k^2} \sum_{i=1}^{n_k} \tau_{k,i}^2, \tag{3.12}$$

where $\text{var}(f(X_{k,i})) = \tau_{k,i}^2$, for $1 \leq i \leq n_k$. To derive an upper bound of the variance Novak uses the variance inequality by Popoviciu (1935) for bounded random variables, to bound the unknown substrata variance $\tau_{k,i}^2 \leq \Delta_{k,i}^2/4$, where $\Delta_{k,i} = f(X_{k,i}) - f(X_{k-1,i})$. Using the monotonicity property of the

function the substrata bound can be further upperbounded as $\Delta_{k,i}^2 \leq \Delta_k^2$, which derives an upper bound of the variance of Novak's estimator,

$$\text{var}(\hat{S}_n^{\text{novak}}(f)) \leq \frac{1}{4} \sum_{k=1}^{m+1} \frac{w_k^2 \Delta_k^2}{n_k} \tag{3.13}$$

Minimizing the upper bound in (3.13) with respect to the constraint budget $\sum_{k=1}^{m+1} n_k = n - m$, where $n - m$ is the leftover budget for allocation in the second step, gives the following allocation rule,

$$n_k^* = \frac{w_k \Delta_k}{\sum_{k=1}^{m+1} w_k \Delta_k} (n - m). \tag{3.14}$$

This allocation rule in (3.14) is proposed by Zhao and Vakili (2008) and it has a direct analogy with the Neyman's allocation rule discussed previously in (3.8). However it turns out that the allocation rule (3.10) used by (Novak, 1992) is also equivalent. In the framework of Novak's sequential algorithm, with $w_k = 1/(m+1)$ for $1 \leq k \leq m+1$ and $\sum_k^{m+1} \Delta_k = 1$, the allocation rule in (3.14) boils down to $n_k^* = \Delta_k \cdot (n - m)$, which is equivalent to allocating the remaining budget $(n - m)$ using weights based on strata bounds. This is almost equivalent to Novak's allocation rule in (3.10), which additionally includes the ceiling function to solve the integer value problem and a smaller allocation budget of $(n - 2m) = m + 1$ instead of $(n - m) = 2m + 1$. This choice of smaller allocation budget is imposed to satisfy the total budget constraint in Novak's algorithm, $\sum_k^{m+1} n_k^* \leq n = 3m + 1$, and it is justified by the following theorem.

Theorem 3.3.1. *Following the construction of the two stage sequential algorithm proposed in Novak (1992), the maximum number of substrata as obtained from* $m + 1$ *equispaced strata is* $2m + 1$.

Proof. Let us recall the sampling budget allocation rule proposed by Novak (1992),

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

For each $k = 1, 2, \ldots, (m+1)$, we have,

$$n_k = (m+1) \cdot \Delta_k + 1 - \varepsilon_k,$$

for some $\varepsilon_k \in [0, 1]$, with $\varepsilon_k = 0 \iff \Delta_k = 0$. Now we have $\varepsilon_k > 0$ for at least one k, because Δ_k can not be zero for all the strata in our framework.

Therefore,

$$\begin{aligned}\sum_{k=1}^{m+1} n_k &= \sum_{k=1}^{m+1}((m+1)\cdot\Delta_k + 1 - \varepsilon_k) \\ &= (m+1)\sum_{k=1}^{m+1}\Delta_k + (m+1) - \sum_{k=1}^{m+1}\varepsilon_k \\ &= 2m + 2 - \sum_{k=1}^{m+1}\varepsilon_k\end{aligned} \tag{3.15}$$

It follows from (3.15) that $\sum_{k=1}^{m+1}\varepsilon_k$ is an integer, and thus $\sum_{k=1}^{m+1}\varepsilon_k \geq 1$ since $\varepsilon_k > 0$ for at least one k. Finally, (3.15) and $\sum_{k=1}^{m+1}\varepsilon_k \geq 1$ yield : $\sum_{k=1}^{m+1} n_k \leq 2m+1$. □

3.3.2.3 Remarks on Novak's algorithm

Novak (1992) proposed this two stage algorithm to prove the rate optimality of the lower bound for sequential randomized methods. It is to be noted that this algorithm as proposed originally, was not sufficiently formalized for a practical application case, since Novak's budget allocation rule (3.10) does not necessarily allocate the full sampling budget all the time. Depending on the function f, the algorithm can generate no additional substrata in the second step, which corresponds to an allocation of only $m+1$ sampling budget in the second step, and that is equivalent to a loss off m sampling budget. This case specifically occurs with functions where all the strata bound differences are equal,

$$\Delta_k = \frac{1}{(m+1)}, \forall\, k = 1, 2, \ldots, (m+1) \tag{3.16}$$

A common example of this type of functions is the linear function $f(x) = x$. Many other examples can be constructed to demonstrate that the budget loss can, at times, be significant, and instead, it could be better utilized for improving the estimates. This serves as the inspiration to create a more practically efficient budget allocation rule.

Theorem 3.3.1 proves that the maximum number of substrata created in the second step is $2m+1$, that is equivalent to full sampling budget allocation. An example function for this case with full budget allocation, can be constructed with the strata bound differences given by,

$$\Delta_k = \begin{cases} \frac{1}{m+1} - m\epsilon & k = k_0, \\ \frac{1}{m+1} + \epsilon & \text{otherwise}, \end{cases} \tag{3.17}$$

for some $\epsilon > 0$ such that $\epsilon < 1/m(m+1)$. Then all the strata are allocated two samples except the k_0-th stratum which has one sample. In other words, the adaptive step generates exactly $2m+1$ substrata, by dividing each but the k_0-th stratum.

3.3.3 . Variance upper bound

3.3.3.1 Novak's upper bound

Using the two stage sequential algorithm Novak (1992) proved that for the class of monotone functions F, the lower bound of order $n^{-3/2}$ is rate optimal. He derived the upper bound of the variance of his proposed algorithm as follows,

$$\begin{aligned}
\mathrm{var}(\hat{S}_n^{\mathrm{novak}}(f)) &\leq \sum_{k=1}^{m+1} \frac{w_k^2}{n_k^2} \sum_{i=1}^{n_k} \frac{1}{4}\Delta_{k,i}^2, && \text{using inequality by Popoviciu (1935)} \\
&\leq \frac{1}{4\cdot(m+1)^4} \sum_{k=1}^{m+1} \sum_{i=1}^{n_k} \frac{\Delta_{k,i}^2}{\Delta_k^2}, && \text{since, } w_k = \frac{1}{m+1} \text{ and } n_k \geq \Delta_k \cdot (m+1) \\
&\leq \frac{2m+1}{4\cdot(m+1)^4}, && \text{using Theorem 3.3.1 and } \Delta_{k,i}^2 \leq \Delta_k^2 \\
&\leq \frac{1}{2\cdot(m+1)^3}.
\end{aligned} \tag{3.18}$$

For a total budget $n = 3m + 1$, the variance upper bound boils down to $(\sqrt{54}/2) \cdot n^{-3/2}$ for the L^1 error.

3.3.3.2 Improvement of the upper bound

Without modifying the original algorithm, the upper bound of variance given by (3.18) can be improved with respect to the constant, by using the monotonicity property of the function under consideration. Moreover it can be shown that the improved upper bound is exactly attained for a function from the class F.

Theorem 3.3.2. *The two stage sequential randomized algorithm proposed by Novak (1992) satisfies*

$$\mathrm{var}(\hat{S}_n^{\mathrm{novak}}(f)) \leq \frac{1}{4\cdot(m+1)^3}. \tag{3.19}$$

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

Proof. 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,

$$\begin{aligned}
\mathrm{var}(\hat{S}_n^{\mathrm{novak}}(f)) &\leq \sum_{k=1}^{m+1} \frac{w_k^2}{n_k^2} \cdot \frac{1}{4} \sum_{i=1}^{n_k} \Delta_{k,i}^2, \quad \text{using inequality by Popoviciu (1935)} \\
&\leq \frac{1}{4} \sum_{k=1}^{m+1} \frac{w_k^2 \Delta_k^2}{n_k^2} \\
&\leq \frac{1}{4 \cdot (m+1)^3}, \qquad \text{since, } w_k = \frac{1}{m+1} \text{ and } n_k \geq \Delta_k \cdot (m+1)
\end{aligned} \tag{3.20}$$

For a total sampling budget $n = 3m+1$, the corresponding L_1 error upper bound turns out to be $(\sqrt{54}/2\sqrt{2}) \cdot n^{-3/2}$, which is less by a factor $\sqrt{2}$, as claimed in Novak (1992).

In order to prove the second part of the claim, we define the following staircase function satisfying (3.16),

$$f(x) = \begin{cases} 0, & 0 \leq x < \frac{0.5}{m+1} \\ \frac{k}{m+1}, & \frac{k-0.5}{m+1} \leq x < \frac{k+0.5}{m+1} \\ 1, & \frac{m+0.5}{m+1} \leq x \leq 1. \end{cases} \text{, for } k = 1, 2, \ldots, m \tag{3.21}$$

For this function, Novak's algorithm does not generate any substrata in the adaptive step, and the allocated sampling budget is $n_k = 1$, for $k = 1, 2, \ldots, (m+1)$. Moreover this particular staircase function is constructed such that for each strata I_k, $\forall k$, there is a jump of size $\Delta_k = 1/(m+1)$, in the function value at the middle of the strata. Therefore inside I_k, the random sample $f(X_{k,1})$ has half of its probability concentrated at each of the two strata bounds, which gives $\tau_k^2 = \Delta_k^2/4$, $\forall k$ to achieve the equality in (3.20). □

3.3.4 . Modifications to Novak's algorithm

3.3.4.1 Budget allocation to strata with zero bound difference

The allocation rule (3.10) proposed by Novak (1992), allocates one sample to each of the strata with $\Delta_k = 0$. We observe that, for such strata, where is function is constant, the integral can be computed deterministically, and this calculation corresponds directly to determining the area of the rectangle formed by that specific stratum. This no longer requires sampling from strata with $\Delta_k = 0$, and the budget saved thereby can be allocated to other strata. We redefine the estimator labeled as $\hat{S}_n^{\mathrm{opt}}(f)$, with this improved strategy of budget allocation,

$$\hat{S}_n^{\mathrm{opt}}(f) = \sum_{\substack{1 \leq k \leq m+1 \\ \Delta_k \neq 0}} w_k \cdot \frac{1}{n_k} \sum_{i=1}^{n_k} f(X_{k,i}) + \sum_{\substack{1 \leq k \leq m+1 \\ \Delta_k = 0}} \frac{f(x_k)}{m+1}. \tag{3.22}$$

For the functions having strata with $\Delta_k = 0$, this improved strategy always allocates more budget than Novak's allocation rule, to the strata with $\Delta_k > 0$.

3.3.4.2 Improved budget allocation rule

As discussed earlier Novak's allocation rule (3.10) not always allocates the full sampling budget, and thus we propose an improved allocation rule to address this issue. Novak (1992) uses a fixed cardinality setup with total budget $n = 3m+1$, out of which m evaluations are used to construct the initial equispaced intervals. From the rest $2m+1$ budget the algorithm tries to allocate $(m+1)$ with the rule $n_k = \lceil \Delta_k \cdot (m+1) \rceil$, such that even if the maximum possible number of the strata are assigned one extra sample due to the ceiling, the total allocation does not exceed $2m+1$. As we have seen before this depends on the function and eventually on Δ_ks and sometimes for special classes of function this rule is not efficient. We thus propose the following allocation rule,

$$n_k = \lceil \Delta_k \cdot \alpha \cdot (n-m) \rceil \tag{3.23}$$

where α is a parameter that maximizes $\sum_{k=1}^{m+1} n_k$ subject to $\sum_{k=1}^{m+1} n_k \leq n-m$. We observe that Novak's allocation rule is a special case of (3.23) with $\alpha = (n-m)/(m+1) = (2m+1)/(m+1) \approx 1/2$. Given that no samples are allocated to the strata with $\Delta_k = 0$, this new rule clearly allocates more (never less) samples than Novak's allocation rule, to the strata with $\Delta_k > 0$, which eventually produces an estimator with a lower variance. However, even when the parameter α is optimally selected, it can lead to scenarios where $\sum_{k=1}^{m+1} n_k < n-m$, implying that a portion of the total sampling budget still remains unutilized. These type of situation arises mainly due to the construction of the allocation rule using a ceiling function, and the amount of unused budget can increase with the number of strata having equal or nearly equal values of Δ_i. In these cases, the remaining budget is allocated sequentially to the strata with decreasing non-zero Δ_i values, using a repetitive framework, continuing until the entire budget has been used up.

To derive the upper bound of the variance with the improved allocation rule in (3.23), we replace $\alpha \cdot (n-m)$ by the maximum possible value of a constant η, such that the allocated budget with the rule $n_k = \lceil \Delta_k \cdot \eta \rceil$, satisfies the constraint $\sum_{k=1}^{m+1} n_k \leq n-m$. This assures the total allocation does not exceed the remaining budget. Following the same arguments as in Theorem 3.3.1, it can be shown that, the total allocated budget using η is always upper bounded by the factor $\eta + m$, and the corresponding lower bound is $\max(m+1, \eta)$.

$$\max(m+1, \eta) \leq \sum_{k=1}^{m+1} n_k \leq \eta + m \tag{3.24}$$

Ensuring the upper bound satisfies the budget constraint $\eta + m \leq n - m$, we have $\eta = n - 2m$. Using the allocation rule $n_k = \lceil \Delta_k \cdot \eta \rceil$ in the upper bound of (3.20) we get,

$$\mathrm{var}(\hat{S}_n^{\mathrm{opt}}(f)) \leq \frac{1}{4 \cdot \eta^2 (m+1)}. \tag{3.25}$$

It can be shown that, this upper bound is also equivalent to $(\sqrt{54}/2\sqrt{2}) \cdot n^{-3/2}$, for the L_1 error, which proves that the new allocation rule does not lose the upper bound guarantee as established in (3.20).

3.3.4.3 Optimal initial number of strata

Novak (1992) uses approximately one third of the total budget to construct the initial equispaced strata. Given the adaptive framework of assigning sampling budget, this might not be an efficient choice from a practical point. Indeed, according to the construction of the algorithm, the initial strata are equispaced, and each of them are assigned at least one sample. Depending on the function, a large number of initial strata with very small values $\Delta_i \approx 0$, will be inefficient since it will assign samples to relatively flat regions. We propose a generalized initial budget of $\lceil n \cdot \gamma \rceil$ with the parameter $\gamma < 0.5$. Note that the constraint $\gamma < 0.5$ ensures the fact that there remains at at least one point as sampling budget for each of the adaptive substrata.

The optimal values of γ is obtained by minimizing the upper bound (3.25), with $m = n\gamma$. Differentiating with respect to γ, we obtain,

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}\gamma} 4(n - 2n\gamma)(n\gamma + 1) = 0 \\ &\implies \gamma^{\text{opt}} = \frac{n-4}{n} \cdot \frac{1}{6} \approx \frac{1}{6} \end{aligned} \tag{3.26}$$

such that the critical value $\gamma^{\text{opt}} = 1/6$ satisfies the second derivative condition for minimizing the upper bound. Using $m \approx n \cdot (1/6)$ in (3.25) we get the L_1 error upper bound is $(\sqrt{54}/4) \cdot n^{-3/2}$. With the improved upper bound in (3.20) and modification proposed respectively on the optimal budget allocation rule, the L_1 upper bound of error is less by a factor 2 compared to the bound claimed in Novak (1992).

3.4 . Numerical experiments

3.4.1 . Methodology

This study focuses on the worst case error of integral approximation methods, and so far have we have studied the upper bounds for approximation error, for different methods. In this section we present a some numerical experiments to monitor the actual empirical mean squared error and the a posteriori upper bounds as obtained adaptively, after splitting the equispaced strata using different budget allocation rules. The integration methods along with the proposed modifications, are tested on different test functions from F, against the original algorithm as proposed by Novak (1992). The numerical benchmark is constructed with two different sets of test functions, described in Table 3.1 and Table 3.2. The first benchmark consists of smooth

functions, which include convex and concave functions with varying gradient parameter as shown in Figure 3.5. Additionally it includes the linear function as a special case. The second benchmark consists of staircase functions, starting from a single step function up to multi-step functions as shown in Figure 3.7. A particular staircase function with p steps is defined based on two sets of parameters, denoted by $\mathcal{A}_p$ and $\mathcal{H}_p$ in Table 3.2, which describe the positions and heights of the steps. The integration methods that are tested, include the following sequential randomized method: Novak's original algorithm called as the *reference*, and two of its modified versions, named *full_alloc* and *full_alloc_init* respectively. The method *full_alloc* implements the new budget allocation scheme proposed in (3.23), which allocates the total sampling budget. The method *full_alloc_init* selects the initial number strata using the optimal rule as proposed in (3.26), in addition to the optimal allocation scheme. The other integration methods that are considered are the simple Monte Carlo method and the nonsequential deterministic trapezoidal rule.

Table 3.1: Test functions from class $F : [0, 1] \to [0, 1]$.

Benchmark	Definition	Parameters
Smooth functions	linear: $f(x) = x$	$-$
	concave: $f(x) = x^{1/p}$	$p = 2, 10, 50$
	convex: $f(x) = x^p$	$p = 2, 20, 50, 100$
Staircase functions	$f(x,p) = \sum_{i=1}^{p} h_i \cdot \mathbb{1}_{\{x-1/\sqrt{a_i}>0\}}$ $\mathcal{A}_p = \{a_1, a_2, \ldots, a_p\}$ $\mathcal{H}_p = \{h_1, h_2, \ldots, h_p\}$	$p = 1, 2, 3, 4, 5$

Table 3.2: Staircase test functions parameters.

Parameter	Set	Values
$p = 1$	$\mathcal{A}_1$	$\{70\}$
	$\mathcal{H}_1$	$\{1\}$
$p = 2$	$\mathcal{A}_2$	$\{70, 20\}$
	$\mathcal{H}_2$	$\{0.5, 0.5\}$
$p = 3$	$\mathcal{A}_3$	$\{70, 20, 10\}$
	$\mathcal{H}_3$	$\{0.33, 0.34, 0.33\}$
$p = 4$	$\mathcal{A}_4$	$\{70, 20, 10, 2\}$
	$\mathcal{H}_4$	$\{0.33, 0.34, 0.18, 0.15\}$
$p = 5$	$\mathcal{A}_5$	$\{70, 20, 10, 2, 1.5\}$
	$\mathcal{H}_5$	$\{0.25, 0.25, 0.25, 0.15, 0.1\}$

Each of the five integration methods are used to approximate the integral $\int_0^1 f(x)\mathrm{d}x$, for $f(x)$ from Table 3.1. The average performance of these meth-

ods are monitored at different level of sampling budget: $20, 40, 60, \ldots, 1000$. For a given sampling budget, the empirical mean squared error for each of the randomized method is used as a metric of comparison, which is estimated using 1000 independent repetitions. For the randomized methods, the a posteriori upper bound of variance given by (3.20) is also plotted on the same figure. For the deterministic trapezoidal rule, the squared error is taken as the measure of performance, and the corresponding maximal squared error is plotted as an upper bound.

Remark 5. It is to be noted that, the quantity in (3.20) is called the a posteriori variance, since it depends on the underlying function and it is computation is based on the allocated budget, inside each strata. This is in no way related to the notion of the a posteriori probability distribution in the Bayesian terminology.

3.4.2 . Results

Figure 3.6 shows the empirical squared L_2 error, or the mean squared error on the smooth benchmark. We observe that, for each of the integration methods, the posterior upper bound of variance, as denoted by the dotted lines, is well respected by the corresponding mean squared error estimates, which are shown in solid lines. Firstly, in terms of the variance upper bounds, the sequential randomized methods have significantly better worst case guarantees compared to the simple Monte Carlo method. The deterministic trapezoidal rule has a lower worst case maximal squared error than the simple Monte Carlo, but it is clearly worse compared to the sequential methods. A closer look into the family of sequential methods reveals that the two proposed methods, namely, *full_alloc* and *full_alloc_init*, produces consistently better posterior upper bounds compared to Novak's original algorithm, called the *reference*. It is to be noted that, as shown in Section 3.3, the two methods *reference* and *full_alloc* has the same worst case upper bound for the variance and only the method *full_alloc_init*, with the optimal initial number of strata, as given in (3.26), has an upper bound, that is better by a factor of two. However, when the posterior upper bounds are considered, it is not evident that the improvement based on optimal initial number of strata, is always beneficial over the *full_alloc* method.

Next, considering the actual mean squared error in estimating the integrals, the naive Monte Carlo method shows consistently worse performance compared to the others. For the linear function, evidently the trapezoidal rule performs the best with almost zero estimation error, which is clearly justified by the construction of the underlying integration rule. Interestingly for nonlinear functions as well, with lower or moderate steepness of the curves, the trapezoidal rule has lower MSE rates, compared to the randomized sequential methods. This is observed with the examples, namely, the concave function

with $p = 2$ and the convex functions with $p = 2$ and 20. However for both type of non linear functions, as the steepness of the curve, i.e., the parameter p increases, the sequential methods show better MSE values. Among the sequential methods, *full_alloc* is consistently better than the *reference* method. The improvement due to *full_alloc_init* is not evident for the concave function, however for the convex case it has the lowest MSE values. These family of convex functions, with steep gradients, hold particular significance in this study, because they closely resemble the shapes of functions generated by the QMRA application in which we are interested in.

Figure 3.8 shows the results corresponding to the staircase benchmark, and we can draw similar conclusions like the preceding benchmark, considering the posterior variance upper bound of the methods. In terms of the estimated MSE, the sequential randomized methods have the lowest errors, followed by the trapezoidal rule and simple Monte Carlo method respectively. The MSE values of the *full_alloc* are generally better than its other sequential counterparts. We observe that for all the randomized methods, the MSE values are very close to their corresponding variance upper bound. This can be attributed to the fact that, most of the initial strata with the staircase functions has zero strata bound difference, which helps in achieving a near equality situation for the inequality used in (3.20). In this specific benchmark featuring non-smooth functions, the fluctuation in MSE rates across various sampling budgets is notable. For such functions the variability is only concentrated in a few specific points, and their position with respect to the integration intervals, is an influencing factor on the integration error. As a result the trapezoidal rule displays pronounced oscillations in the squared error values.

To summarize, the sequential randomized methods show clear benefits over the simple Monte Carlo method for this class of monotone functions. The trapezoidal rule can be advantageous in certain cases; however, as the increasing steepness of smooth non-linear functions or with non-smooth staircase functions, it is outperformed by sequential methods. Within the set of sequential methods, the *full_alloc* method consistently demonstrates improvement over the *reference* method for all the considered test problems.

3.5 . Application to QMRA simulator

3.5.1 . Problem description

This chapter is focused on finding a suitable cost effective integration method, for the special type of functions that are encountered in the domain of QMRA modeling. To evaluate the performance of the integration methods studied in the previous sections, this section will present a relatively simpler version of the original problem of integration in the QMRA model under consideration.

As described in the introduction of this chapter, the aim is to compute ex-

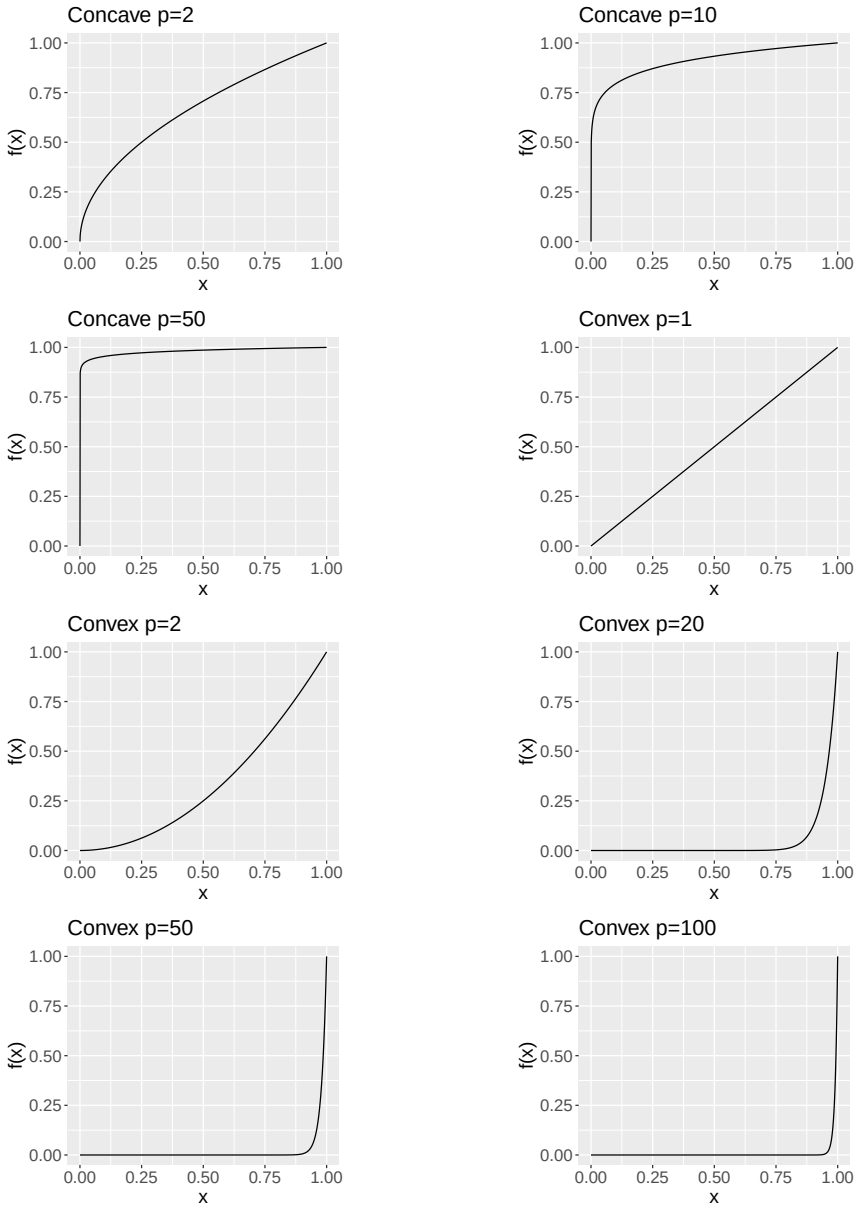


Figure 3.5: Smooth test functions: concave, convex and linear.

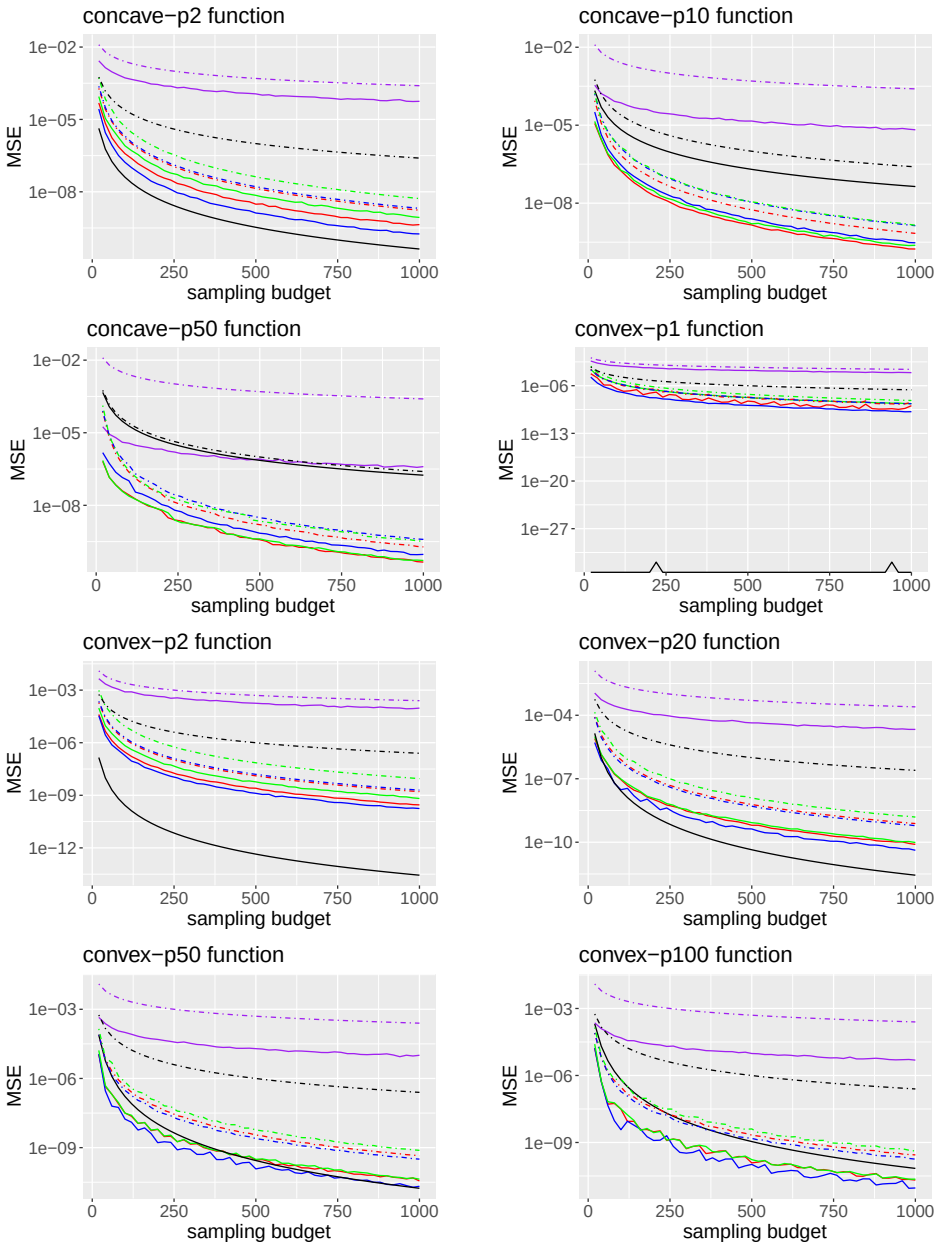


Figure 3.6: Empirical MSE (solid lines) and corresponding upper bounds (dash dot lines) for the smooth benchmark, as obtained with the methods: Trapezoidal rule (black), simple Monte Carlo (purple), *full_alloc* (red), *full_alloc_init* (blue) and *reference* (green).

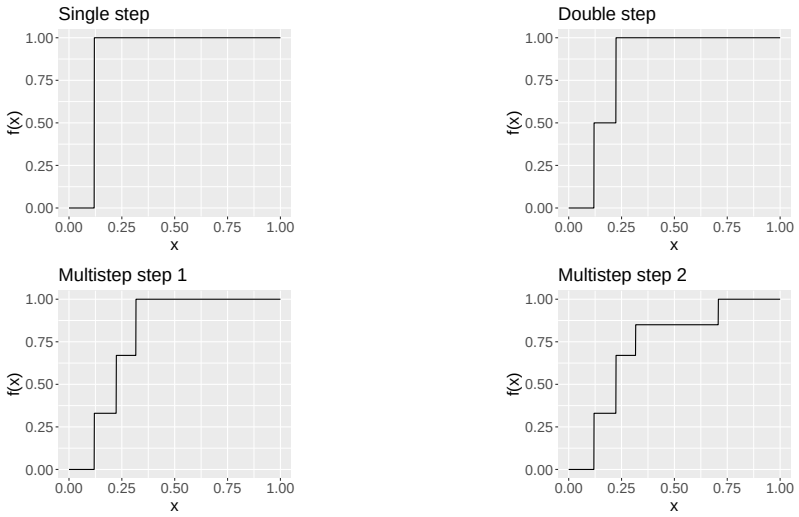


Figure 3.7: Staircase test functions

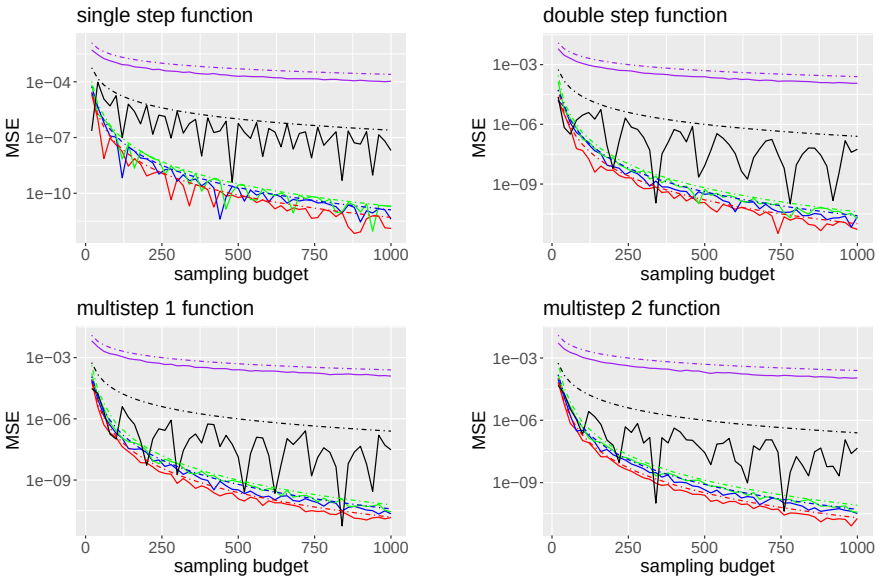


Figure 3.8: Empirical MSE (solid lines) and corresponding upper bounds (dash dot lines) for the staircase benchmark, as obtained with the methods: Trapezoidal rule (black), simple Monte Carlo (purple), *full_alloc* (red), *full_alloc_init* (blue) and *reference* (green).

pectations of two monotone and bounded outputs of the QMRA simulator, in order to estimate the ultimate quantities of interest, namely, the average risk and the average probability of rejecting a batch. Firstly the risk of food-borne illness $R_x^{\text{batch}}(\Xi_x)$, corresponding to the pathogen $x \in \{\text{MPS-STEC}, \textit{Listeria monocytogenes}, \textit{Salmonella}\}$ by consuming a cheese portion from a particular batch of cheese, and secondly the probability of rejecting a batch of cheese $P^{\text{batch}}(\Xi)$, due to contamination by any of the three pathogens. These two outputs are functions of the vector of stochastic internal variables $\Xi = (\Xi_x)_x$ of the QMRA model (see Section 2.2.6.1 in Section 2.2.6.1), and the goal is to compute their expectations with respect to Ξ. As demonstrated in Figure 3.1 both the outputs are partially monotonically increasing with respect to the initial concentrations of the respective pathogens $Y_x^{\text{milk}} \in \Xi_x$. Hence this section simplifies the original problem by introducing a first hypothesis of considering Y_x^{milk} as the integrating variable in both cases.

It is to be noted that, the multi-pathogen QMRA framework involves two different types of integration problems. The expectation of batch risk comprises of three separate uni-variate problems of integrating $R_x^{\text{batch}}(\Xi_x)$ with respect to Y_x^{milk}, corresponding to three different pathogens x under consideration. The expectation of probability of rejection is a multivariate integration problem of $P^{\text{batch}}(\Xi)$ with respect to each of the three initial pathogen concentrations Y_x^{milk}. Given the scope of this study, this section is restricted only to the problem of uni-variate integration, and hence the problem is further simplified by considering only the outputs corresponding to the single pathogen model based on MPS-STEC. In other words, we consider two functions $g_1 \coloneqq R^{\text{batch}}_{\text{MPS}-\text{STEC}}(\Xi_{\text{MPS}-\text{STEC}})$ and $g_2 \coloneqq P^{\text{batch}}(\Xi_{\text{MPS}-\text{STEC}})$, and the aim is to compute their expectations with respect to $Y^{\text{milk}}_{\text{MPS}-\text{STEC}}$.

3.5.2 . Test case construction

In practice, the QMRA simulator integration problem is more complicated, mainly due to two bottlenecks. The first bottleneck is the presence of the other stochastic internal variables apart from $Y^{\text{milk}}_{\text{MPS}-\text{STEC}}$ in $\Xi_{\text{MPS}-\text{STEC}}$. More precisely these are, 1) the milk storage duration d^{storage}, a Triangular$(1, 12, 40)$ distributed random variable, 2) the milk storage temperature T^{storage} which is an $\text{Uniform}(4, 6)$ distributed random variable, and 3) the time of consumption t^{consum}, distributed as Triangular$(22, 30, 60)$. In order to obtain partial monotonicity of the outputs with respect to $Y^{\text{milk}}_{\text{MPS}-\text{STEC}}$, as shown in Figure 3.1, these stochastic internal variables can either be fixed or can be averaged over their respective domains. The batch risk $R^{\text{batch}}_{\text{MPS}-\text{STEC}}$ is integrated with respect to t^{consum}, by approximating $\mathbb{E}_{t^{\text{consum}}}[R^{\text{batch}}_{\text{MPS}-\text{STEC}}(t^{\text{consum}})]$ using a piecewise constant function on a regular grid $\{t_1^{\text{consum}}, t_2^{\text{consum}}, \ldots, t_n^{\text{consum}}\}$ of size n, that spans the support of the triangular distribution of the variable t^{consum}. The other two stochastic parameters, namely, d^{storage} and T^{storage}, serves as

inputs of a computationally expensive cheese processing module inside the QMRA simulator, as a result, numerically integrating the outputs with respect to these parameters includes several runs of the cheese module, which is computationally expensive. Hence the parameter d^{storage} is kept fixed at its mode value 12 hours, and T^{storage} is fixed at its mean 5 degree Celsius.

The second bottleneck in the practical aspect is the unavailability of knowledge about the distribution of the integrating variable. The sequential and nonsequential methods and the associated results as studied in the previous sections, are applicable when the integral is over the domain $[0, 1]$ and the integrating variable is uniformly distributed in the unit interval. As indicated in the introduction, to transform the integration problem in $[0, 1] \to [0, 1]$, we use the transformation of integrating variable $x = F_Y(y)$, which requires the knowledge about the distribution of Y. In absence of the functional form of the original distribution of the integrating variable, the corresponding empirical estimates of the quantiles can be used. However these quantile estimates include an estimation error which are additionally incorporated to the estimation error of the integration method. Depending on the computational cost associated with simulating the integration variable, this error has the potential to cancel the theoretical advantages gained from employing the proposed integration methods. This bottleneck is addressed using distributional assumption on the initial concentration of MPS-STEC. A LogNormal distribution is used to fit the simulated values of the initial MPS-STEC concentration, by equating the first and second order moments. The fitted distribution is plotted against the original simulated values in Figure 3.9, along with a QQ plot between the original and the fitted distribution. While it seems from the QQ plot that the true distribution may have slightly heavier tails compared to the approximated one, we still consider it a reasonable approximation to work with.

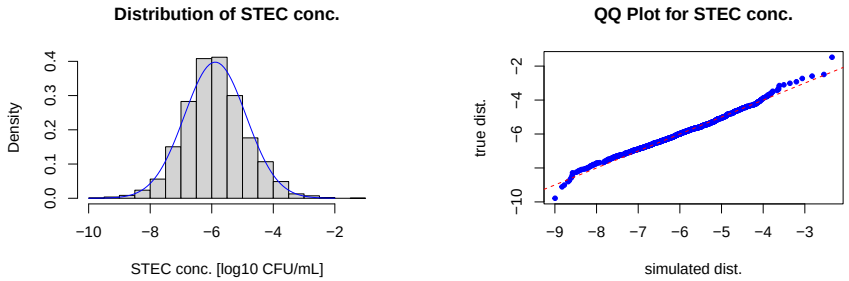


Figure 3.9: Approximated distribution with mean -5.9 and standard deviation 1.0 with the corresponding QQ plot, for the integrating variable in $\log 10$ scale.

3.5.3 . Methodology

Once the probability distribution of the integrating variable is known, the transformation of variables can be used to get suitable integrating problems, on the transformed functions of g_1 and g_2. To evaluate the performance of the above studied integration methods on this specific problems, we take one further step to approximate these transformed functions by simple non-linear functions. The integration methods are tested on these non-linear functions, for which the true value of the integral is analytically computable, which in turns allows to compute the performance metrics similarly as in the previous numerical benchmarks. We consider respectively a convex function and a sigmoid function, as surrogates models, to approximate the two integrands. The fitting is done using the *nls()* function in R, used for nonlinear least squares regression. It fits the nonlinear surrogate model by optimizing the model parameters, that minimize the sum of squared differences between the observed and the predicted values. The non-linear surrogate models with the corresponding estimated model parameters are listed in Table 3.3. Figure 3.10 plots the transformed QMRA functions $f_i := g_i \circ F^{-1}_{Y^{\text{milk}}_{\text{MPS}-\text{STEC}}}$, where $f_i : [0,1] \to [0,1]$, for $i = 1, 2$ and $F_{Y^{\text{milk}}_{\text{MPS}-\text{STEC}}}$ is the approximated distribution function of the integrating variable $Y^{\text{milk}}_{\text{MPS}-\text{STEC}}$.

Table 3.3: QMRA test functions non-linear approximations.

Function	Non-liner surrogate	Parameters
g_1	convex: $f(x) = x^p$	$p = 328475$
g_2	sigmoid: $f(x) = \frac{1}{1+\exp(-\alpha\cdot(x-\beta))}$	$\alpha = 16.4$, $\beta = 0.889$

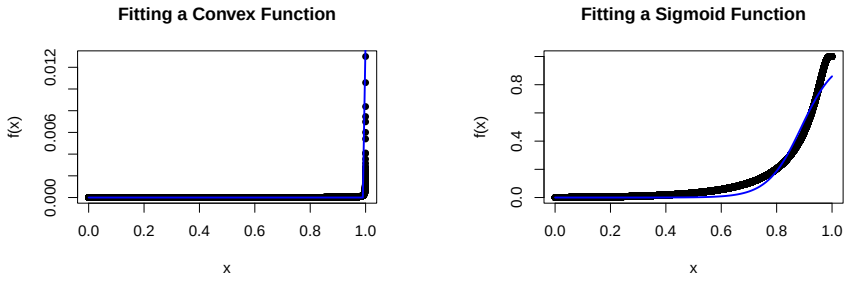


Figure 3.10: Transformed QMRA functions f_1 and f_2 plotted against the transformed integrating variable $x = F_{Y^{\text{milk}}_{\text{MPS}-\text{STEC}}}(y^{\text{milk}}_{\text{MPS}-\text{STEC}})$.

A similar framework as described in Section 3.4 is followed to evaluate the QMRA benchmark. The same set of integration methods and their corresponding worst case upper bounds are compared, at different levels of sampling budget of interval of 100, with a maximum total budget of 10000 and a

minimum budget 20, and the mean squared errors are estimated using 1000 independent iterations for the randomized methods.

3.5.4 . Results

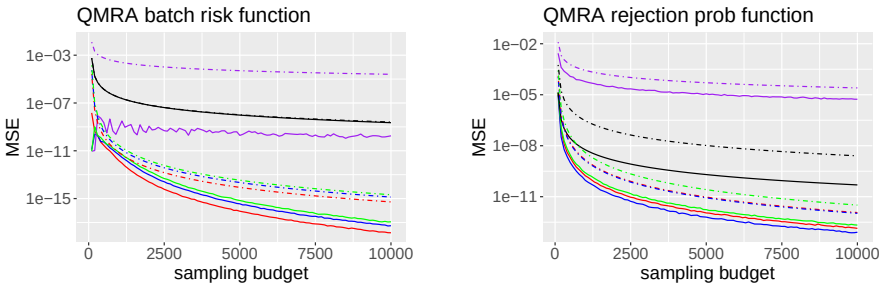


Figure 3.11: Empirical MSE (solid lines) and corresponding upper bounds (dash dot lines) for the QMRA functions, as obtained with the methods: Trapezoidal rule (black), simple Monte Carlo (purple), *full_alloc* (red), *full_alloc_init* (blue) and *reference* (green).

Figure 3.11 monitors the empirical squared L_2 error, or the mean squared error on the QMRA benchmark, along with the worst case upper bounds. The upper bounds of the variances show similar behavior in compared to the previous benchmarks, for both of the functions. The proposed variants of sequential methods provide significantly better worst case guarantees compared to the simple Monte Carlo method and the trapezoidal rule. As observed in the earlier experiments also, the improvement of the *full_alloc_init* method is not evident over the *full_alloc* method, however both of them produces better worst case guarantees than Novak's method. Concerning the MSE values as well, the proposed sequential methods perform better than the others. For the sigmoid function, the trapezoidal rule is still better than the simple Monte Carlo method, but for the highly steep convex function it almost reaches the worst case bound. From the results it is evident that for step-like functions, use of sequential randomized methods can bring significant improvement in the results.

3.6 . Discussion and perspectives

The motivation of this study was to leverage the boundedness and monotonicity properties of the QMRA simulator outputs, in order to reduce the computational cost of estimating the average quantities of interest. Considering the numerical benchmarks presented so far in this chapter, we can safely say that the sequential randomized methods are more efficient and cost effective than the simple Monte Carlo integration method. We have also seen

that the deterministic trapezoidal rule can be useful in some cases, mostly when the function is very smooth. But at the same time it can be highly inefficient for step-like or staircase functions, where the functional variance is concentrated locally, for example the functions considered in the QMRA application case. The proposed sequential methods improves over the trapezoidal rule for these type of functions. In a fixed sampling budget framework, our results show that the proposed sequential methods are more efficient in terms of budget allocation and they provide better worst case guarantees on the integration errors, over Novak's original algorithm. The numerical benchmarks confirm that the actual L_2 error is also lower with the proposed modifications.

These sequential methods can be further improved taking into account the particular type of functions we are interested in our application case. For near step like functions as shown in Figure 3.10, where almost all the variation is supposed to be concentrated in one particular point, a "greedy" strategy can be much more efficient. Instead of considering equispaced strata in the first stage of the sequential algorithm, the "greedy" strategy splits the strata using a dichotomy based on strata bound differences. Consequently, the algorithm improves convergence by identifying the region with high variation and allocating more sampling budget there compared to the flat regions. Preliminary numerical experiments support this conjecture and a potential future direction of this work would be to explore this special type of algorithms and provide worst case guarantees.

In Section 3.5, we have demonstrated the application of the proposed integration methods on a simplified version of the original integration problem associated with the construction of the QMRA simulator. In practice the integration problem is more complicated and has several constraints as already discussed. Fixing the stochastic parameters to certain values is a simple and easy to implement solution to this problem, but ignoring the underlying uncertainty might have consequences on the robustness of the QMRA model. On the other hand this problem can also be addressed by averaging out the effects of these stochastic parameters by integrating the final quantities with respect to them, but in some cases it could be computationally expensive enough, to offset the benefits of the integration methods. For the pathogens STEC and *Salmonella*, the batch risk can be analytically integrated with respect to the consumption time, but for *Listeria monocytogenes*, this step becomes computationally expensive as it includes solving an additional series of ordinary differential equations for each time point in the grid, due to a secondary growth step of this particular pathogen before the consumption. Hence, in practice there exists a trade-off between reducing the cost of actual integration and reducing the original integration problem into a suitable format of a bounded, monotone function. In this study, we have only focused on the construction of a uni-variate integration method but in practice the multi-

pathogen QMRA model produces multivariate integration problem. Directions for future work include extending our results to multivariate integration problems with partial monotonicity (see, e.g., McKay et al., 1979, Section 2.1).

Another major perspective of this work is to study the applicability and usefulness of the proposed integration methods, when the distribution F_Y of the integrating variable is analytically unknown, and it is only possible to sample from it at a reasonable cost. In such cases the transformation of the integrating variable depends on the empirical estimates of quantiles from simulated values. Estimates with certain level of accuracy demands higher number of samples, which can become computationally expensive, depending on the simulation cost of the integration variable. For our application case except the pathogen *Listeria monocytogenes*, the distribution for initial concentration Y_x^{milk} of the other two pathogens are unknown, and the simulation has a non negligible computational cost. Future research directions involve exploring the balance between the budget needed to obtain sufficiently accurate quantile estimates and the potential benefits achieved through the utilization of monotonicity and boundedness properties. In terms of the QMRA simulator, there exists a trade-off between the cost of sampling from the farm module and the cheese processing module, which determines the efficiency and applicability of such sequential integration methods.

4 - Multiobjective Optimization

4.1 . Introduction

Multiobjective optimization (MOO) finds application in various industrial domains, aiding in system design and policy-making across a large spectrum of contexts and applications. MOO entails optimizing multiple objectives, or quantities of interest (QoI), by identifying input parameters that establish a balanced compromise among all evaluated objectives. Direct, simultaneous optimization of all objectives is typically infeasible due to the inevitable trade-offs among them.

In the context of food safety and quantitative microbiological risk assessment (QMRA), as documented, e.g., by Basak et al. (2024) and Perrin et al. (2014), policymakers aim to determine optimal process intervention parameters that minimize both the risk of food-borne illness and the cost of implementing intervention measures. A conflict arises between the two objectives, as implementing monitoring and intervention processes reduces the risk of illness but simultaneously increases production costs, and vice versa.

In raw milk cheese production, for example, strategies to minimize bacterial contamination might include performing both pre-harvest and post-harvest tests on farm milk and cheese sample batches, respectively. Reducing the costs related to these intervention steps impacts the hygiene and quality of farm milk and the subsequent cheese, thereby increasing the risk of food-borne illnesses. Policymakers need to identify input parameters for these intervention processes that optimally balance risk and cost factors.

Although a single optimal solution for selecting intervention parameters does not exist in practice, it is viable to identify a set of optimal solutions, denoted as $\mathcal{P}$, known as the Pareto set. The Pareto set comprises optimal solutions, within which no solution can improve one objective without adversely affecting another. This study is confined to a finite input space, with the objective of accurately estimating both the Pareto set $\mathcal{P}$ and its image, termed the Pareto front.

4.1.1 . Problem Formulation

We address the problem of multi-objective optimization for a computationally expensive and stochastic simulator, which permits only a limited number of evaluations and produces noisy outputs. Let $\mathbb{X}$ represent a finite subset of $\mathbb{R}^d$, which will be referred to as the search space, and let $f = (f_j)_{1 \leq j \leq q}$ be a vector of q functions $f_j : \mathbb{X} \to \mathbb{R}$, or objectives, to minimize. The observable data consist of a sequence of noisy evaluations at points $x^1, x^2, \ldots \in \mathbb{X}$. The observations are modeled by random variables $Z_j^i = f_j(x^i) + \varepsilon_j^i$, for $1 \leq j \leq q$,

$i = 1, 2, \ldots$, where the ε_j^is denote random noise. The objective is to determine an estimate of the solution set to the problem

$$\min_{\mathbb{X}} f_1, f_2, \ldots, f_q \,. \tag{4.1}$$

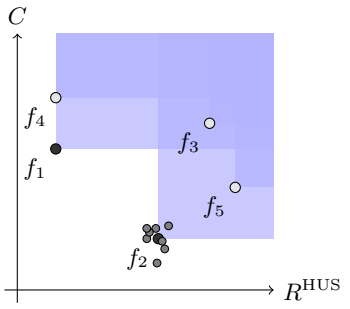
In scenarios where potential trade-offs exist between the q objective functions, the simultaneous minimization is guided by the Pareto domination rule ($\prec$). Consider two input points $x, x' \in \mathbb{R}^d$ with corresponding noise-free function values $z = f(x)$ and $z' = f(x')$. In the context of Pareto domination, $f(x) \prec f(x')$ signifies that x is deemed to dominate x', that is when $z_j \leq z'_j$ for all j, and at least one of the inequalities is strict:

$$f(x) \prec f(x') \iff \begin{cases} \forall j \leq q, & z_j \leq z'_j \,, \\ \exists j \leq q, & z_j < z'_j. \end{cases} \tag{4.2}$$

The set $\mathcal{P}$ of all Pareto optimal points is called the Pareto set:

$$\mathcal{P} = \{x \in \mathbb{X} : \nexists x' \in \mathbb{X} \text{ s.t. } f(x') \prec f(x)\} \,. \tag{4.3}$$

Figure 4.1 illustrates the concept of Pareto optimality with a bi-objective example and two conflicting objectives: the risk of Haemolytic Uremic Syndrome (R^{HUS}) and the cost of intervention (C), derived from the cheese production case study. Each point in the objective space is associated with a shaded rectangle that indicates the corresponding dominated region. In essence, any other point that falls inside the dominated region will present a higher risk or cost compared to the referenced point and, therefore, will be considered a non-optimal choice in the sense of Pareto domination.



f_3, f_4 & f_5 dominated by f_1 and f_2

Figure 4.1: Pareto front constituted with the points z^1 and z^2, that dominate the other points z^3, z^4 and z^5. Shaded rectangles to the top right of every point shows the dominated region by that point. Noisy observations are shown around the point z^2.

In Figure 4.1, the points $z^1 = f(x^1)$ and $z^2 = f(x^2)$ are Pareto optimal since they do not fall in any of the dominated regions and all the other points

namely $z^3 = f(x^3)$, $z^4 = f(x^4)$ and $z^5 = f(x^5)$ fall within their dominated region. In our framework we deal with a stochastic simulator that produces noisy outputs as shown around the point z^2 in Figure 4.1. Identical inputs can produce different outcomes across two evaluations.

The aim is to find a good estimate of the set $\mathcal{P}$ and the Pareto front $\mathcal{F} = f(\mathcal{P})$, using a sequence of evaluation points $\mathbb{X}_n = \{X^1, X^2, \ldots, X^n\}$ and the corresponding noisy evaluations. The estimate of $\mathcal{P}$ will be denoted by $\widehat{\mathcal{P}_n}$, and the estimate of $\mathcal{F}$ will be denoted by $\widehat{\mathcal{F}_n}$.

We also consider that the objectives do not have an analytical form and we have no information on derivatives. In that sense, the objective functions are considered as "black-box" functions. In the literature (see, e.g., Hunter et al., 2019), this type of problem is sometimes called multiobjective simulation optimization (MOSO).

In our example using the QMRA simulator, the objectives of interest—namely, the risk of illness and the cost of intervention—are defined as functions of the expectations of the simulator outputs. The noisy outputs are handled by simulating a large number of batches and computing their expectations, which, in turn, increases the computational cost. Problems exhibiting the aforementioned characteristics can be tackled using Bayesian optimization algorithms, discussed in the following subsection.

4.1.2 . Literature review on Bayesian optimization

The main concept within the Bayesian optimization framework involves modeling each objective function f_j, for $j = 1, 2, \ldots, q$, using a random process, or more specifically, a Gaussian process ξ_j. The use of Gaussian processes as a surrogate model, also known as kriging, is widely adopted in the literature of design and analysis of computer experiments (see, e.g., O'Hagan, 1992, Sacks et al., 1989, Santner et al., 2003) and in the domain of machine learning (see, e.g., Rasmussen and Williams, 2006). Using Gaussian process priors makes it possible to construct a surrogate model for the computationally intensive simulator by obtaining the posteriors of the ξ_js given the evaluations $Z^1, Z^2, \ldots$, which are then used to sequentially construct the evaluations points $\mathbb{X}_n$, and estimates of $\mathcal{P}$ and $\mathcal{F}$. Bayesian optimization is an iterative process, where the Gaussian process surrogate model is used to represent prior beliefs about the unknown objective functions f_j, and this model is updated as new data points are collected, incorporating the Bayesian principle of updating beliefs with evidence. This iterative process is governed by a sampling criterion also known as the acquisition function, which is optimized at each iteration, in order to select a new data point.

In the literature, various Bayesian multi-objective optimization algorithms have been proposed, but only a few are applicable within a stochastic framework (see, e.g., Gonzalez and Nieuwenhuyse, 2020). We will distinguish four

types of approaches.

Entropy-based approaches For the problem of stochastic single-objective optimization, one of the earliest contributions to the noisy evaluation setup is by Vazquez et al. (2008), who compared the empirical convergence rates of several Bayesian optimization algorithms from the literature, in presence of additive noise. In this work, the authors mainly studied the Expected Improvement (EI) sampling criterion (Jones et al., 1998, Močkus, 1975) and the Informational Approach to Global Optimization (IAGO) method, proposed by Villemonteix et al. (2007); both of which were originally proposed in the framework of noise-free evaluations. The IAGO algorithm is grounded in the information theoretic approach, which minimizes the posterior Shannon entropy of the minimizer, to select the next evaluation point in the input space. This algorithm is based on the principle of Stepwise Uncertainty Reduction (SUR) (Geman and Jedynak, 1996, Vazquez and Bect, 2009, Vazquez and Martinez, 2006), a well established framework in the domain of Bayesian sequential design of computer experiments.

Other entropy based algorithms proposed in the machine learning literature include the Predictive Entropy Search (PES) approach (Hernández-Lobato et al., 2014), which reformulates the entropy reduction based on mutual information given the evaluations.

Based on PES, Hernandez-Lobato et al. (2016) proposed the PESMO algorithm, which can be used to jointly estimate the Pareto set and front in the case of stochastic multi-objective problems. PESMO is based on a set of approximations for the computation of the entropy based criterion. This has a high implementation complexity and a non-negligible execution cost, since it relies at each iteration on conditional simulations of the GP models and an expectation-propagation step. To further address the problem of approximations of the entropy and reduce the computational complexity, Wang and Jegelka (2017) proposed a criterion in the single objective framework, called the Max-value entropy search (MES). Later Belakaria et al. (2020) proposed an multiobjective algorithm called MESMO, based on the same principle of maximizing the information gain about the optimal Pareto front given the evaluations. However, these entropy-based algorithms still remains computationally intensive.

Knowledge Gradient Frazier et al. (2009) proposed an approach for Bayesian stochastic single-objective optimization called the Knowledge-gradient sampling criterion, which was derived by revisiting the assumptions of the EI sampling criterion. Subsequently, using this Knowledge-gradient approach, Astudillo and Frazier (2017) proposed an algorithm for the stochastic multi-objective setup.

GP-UCB Another existing algorithm, based on the Gaussian Process Upper Confidence Bound rule (GP-UCB) (see, e.g., Srinivas et al., 2010), named Pareto Active Learning (PAL), was proposed by Zuluaga et al. (2013) and later extended by Barracosa et al. (2021) to Pareto Active Learning for Stochastic simulators (PALS) to address noisy evaluations. The Bayesian ranking and selection literature, akin to Bayesian optimization, considers the input space to be discrete and finite and assumes that the observations are necessarily noisy.

Ranking and Selection The primary objective is to optimally allocate the simulation replications to the input design points so that the Pareto set can be estimated with high confidence. In this context, Lee et al. (2010) proposed an algorithm for Multi-objective Optimal Computing Budget Allocation (MOCBA), which assumes independent Gaussian distributed objectives and secures the optimal allocation by minimizing the asymptotic upper bounds of misclassification error of the design points. This results in a constrained optimization problem solved by the Lagrangian method. Additional algorithms include SCORE by Pasupathy et al. (2014) and M-MOCBA proposed by Branke and Zhang (2015). Recently, Rojas Gonzalez et al. (2020) introduced an approach that incorporates a mixture of Bayesian optimization and ranking and selection approaches, dubbed SK-MOCBA.

4.1.3 . Outline

The main contribution of this chapter is a Bayesian stochastic multi-objective optimization (BSMOO) algorithm based on a weighted integrated mean squared error (w-IMSE) sampling criterion for estimating the Pareto front.

This chapter is organized as follows. Section 4.2 gives a detailed explanation of the problem of multi-objective optimization for stochastic simulators using Bayesian optimization algorithms. Section 4.3 revisits the principle of Maximal Uncertainty Sampling (MUS) and provides an overview of algorithms that employ this principle, also including experimental results discussing the performances of algorithms built upon this principle. Section 4.4 introduces the proposed algorithm, based on a w-IMSE sampling criterion for selecting new points. Section 4.3.4 and Section 4.4.2 present the numerical experiments and their results, and Section 4.5 illustrates the application case of optimizing the QMRA simulator. Finally Section 4.6 discusses the perspectives of this work.

4.2 . Stochastic Bayesian multi-objective optimization

This section describes the strategy of Bayesian optimization for computationally expensive simulators with noisy outputs. In a multi-objective opti-

mization framework, the objective is to identify solutions that provide a compromise among multiple objectives, as determined by the Pareto domination rule. Thus, the task primarily involves obtaining a reliable estimation of the Pareto set $\mathcal{P}$ and its corresponding Pareto Front $\mathcal{F}$. Due to the absence of an analytical form of the objectives and the unavailability of gradient information, this estimation is conducted using a restricted set of evaluation points $\mathbb{X}_n$. The set of evaluation points is selected adaptively, wherein the experimenter sequentially observes and decides at each stage which new point should be evaluated, based on the information collected from the previous evaluations.

4.2.1 . Bayesian optimization framework

Recall the notations from the previous section: $f = (f_j)_{1 \leq j \leq q}$ denote the q objectives subject to minimization, and the random observations are denoted by $Z_j^i = f_j(x^i) + \varepsilon_j^i$, for $1 \leq j \leq q$, and $i = 1, 2, \ldots$, where the ε_j^is are mutually independent Gaussian distributed random variables, $\varepsilon_j^i \sim \mathcal{N}(0, \tau_j^2)$.

In the Bayesian optimization framework, each objective f_j is assigned a prior probability distribution, commonly defined by a Gaussian process, within the following model:

$$\begin{cases} Z_j^i \mid \xi_j, \tau_j^2 \sim \mathcal{N}(\xi_j(x^i), \tau_j^2)\,, \quad j = 1, \ldots, q, \text{ and } i = 1, 2, \ldots \\ \xi_j \mid m_j, k_j \sim \mathrm{GP}(m_j, k_j)\,, \quad j = 1, \ldots, q, \end{cases} \tag{4.4}$$

where the ξ_js are independent random Gaussian processes indexed by the elements of $\mathbb{X}$, with mean functions m_j and covariance functions k_j, and the elements Z_j^i are conditionally independent given ξ_j and the hyperparameters (τ_j^2, m_j, k_j for $1 \leq j \leq q$).

Using Gaussian processes facilitates the computation of the posterior distributions of the models ξ_j, conditional on the observations Z_j^i, through the resolution of systems of linear equations (see, e.g., Chiles and Delfiner, 2012, Cressie, 1993, Rasmussen and Williams, 2006, Santner et al., 2003, Stein, 1999, Welch et al., 1992).

The sequence of evaluation points $(X^1, X^2, \ldots)$, also known as the sampling strategy, is determined adaptively using an iterative procedure as described in Algorithm 1.

At each step n, a Gaussian process model ξ_n is constructed using the data $I_n = \{(X^1, Z^1), \ldots, (X^n, Z^n)\}$, starting from an initial design I_{n_0} of n_0 points, typically obtained from random sampling methods.

In Bayesian optimization, the construction of the GP models based on the design I_{n_0} serves as the initial step, followed by an iterative procedure, wherein each iteration, the Gaussian process model is updated with the observations accumulated thus far. Each step computes the posteriors for the GP models, which are used to compute a sampling criterion, denoted by J_n :

Algorithm 1 BSMOO framework

Place Gaussian process priors on f ▷ (surrogate model)
Evaluate f at n_0 points ▷ (initial design)
$n \leftarrow n_0$ and budget $\leftarrow n_0 \times k$
while budget > 0 **do**
  Update : GP posterior
  Compute : Sampling criterion $J_n(x)$
  Optimize : $x_{n+1} = \text{argmax}_{x \in \mathbb{X}} J_n(x)$ ▷ (new observation)
  Evaluate : f at x_{n+1}
  $n \leftarrow n + 1$ and budget $\leftarrow$ budget $+ k$ ▷ (increment budget)
end while
Estimate $\hat{\mathcal{P}}_n$ and $\hat{\mathcal{F}}_n$ from the GP posterior means

$\mathbb{X} \to \mathbb{R}$. The sampling criterion is optimized (either minimized or maximized) over the search space $\mathbb{X}$ to determine the point to be sampled next. At each sample point the objective functions are evaluated with a batch size k, (described later in Section 4.2.3). At iteration n the Pareto front $\hat{\mathcal{F}}_n$ and the Pareto set $\hat{\mathcal{P}}_n$ are estimated using the posterior means of the Gaussian Processes $\xi_n = (\xi_{j,n})_{1 \leq j \leq q}$ conditional on I_n.

The construction of J_n depends on the type of optimization problem and how the user wants to measure the performance of the optimization algorithm. For instance, consider the case of the expected improvement sampling criterion (Močkus, 1975) in the context of single-objective, noise-free optimization. The expected improvement is a common sampling criterion that seeks the input location which, in expectation, most improves upon the best-known objective value. It is easy to comprehend and use in a noise-free, single-objective scenario, but difficult to generalize in a multi-objective stochastic optimization framework due to the intricate nature of handling multiple conflicting objectives and noise in the observations.

In the subsequent sections, we will explore new sampling criteria derived from two classical foundational ideas: maximal uncertainty sampling (MUS) and stepwise uncertainty reduction (SUR). MUS, as detailed in Section 4.3, is driven by the intent to sample the point in the input space that is most uncertain, with uncertainty being understood in the context of the knowledge about the Pareto front and the Pareto set. Conversely, SUR, introduced in Section 4.4, formulates the sampling criterion to quantify and minimize the residual uncertainty in a look-ahead setup, with the aim of identifying the next evaluation point with minimized uncertainty regarding the aforementioned Pareto entities.

In the following section, we will present several metrics that specifically address the approximation of the Pareto optimal solutions. These measures,

framed within the context of Bayesian optimization, seek to assess the imprecision inherent in approximations of the Pareto front and the Pareto set during the optimization process. The intention is to equip the optimization procedure with metrics to effectively guiding the sampling towards regions that are not only promising in terms of objective function values but also in the minimization of the uncertainty associated with the approximated Pareto solutions.

4.2.2 . Performance metrics

The performance of optimization algorithms is assessed using performance metrics, which may focus either on the quality of estimation of the Pareto set $\mathcal{P}$ or the Pareto front $\mathcal{F}$. These metrics are computed and averaged over several experiments to estimate the average performances of the algorithms.

Volume of symmetric difference. In this chapter, the main performance metric will be the volume of symmetric difference, which focuses on the estimation of the Pareto front $\mathcal{F}$.

Computation of these performance metrics requires a reference Pareto set / front (which might be the true known or hypothesized Pareto set / front) in order to make a comparison. Let $\mathcal{F}^*$ and $\hat{\mathcal{F}}$ be a reference Pareto front and an estimated Pareto front.

Let also R denote a reference point in the objective space, which is user-defined. In practice, the reference point can be defined as an upper-bound of the objectives.

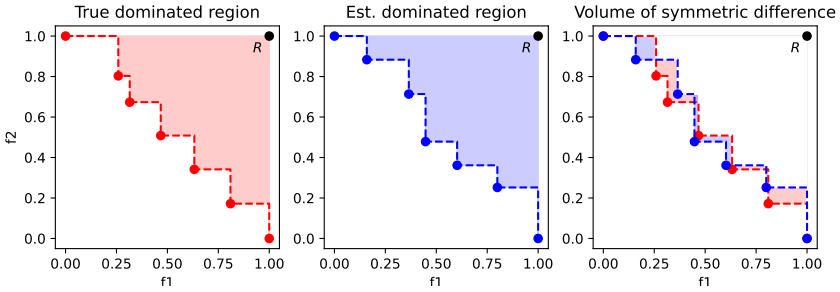


Figure 4.2: Figures from left to right respectively show the dominated area (in red) by the reference Pareto front $\mathcal{F}^*$, the area (in blue) dominated by the estimated Pareto front $\hat{\mathcal{F}}$ and the symmetric difference as the shaded region.

The region dominated by a Pareto front $\mathcal{F}$ with respect to a reference point R is the set of all the points that simultaneously dominate R and are dominated by a member of $\mathcal{F}$:

$$D(\mathcal{F}) = \bigcup_{y^* \in \mathcal{F}} \{y \in \mathbb{R}^q : y^* \prec y \text{ and } y \prec R\} \tag{4.5}$$

Let $V(\mathcal{F})$ denote the hyper-volume of the region dominated by $\mathcal{F}$ with respect to R. Then, the volume of the symmetric difference of the dominated regions corresponding to $\mathcal{F}^*$ and $\hat{\mathcal{F}}$ is defined as

$$V(\mathcal{F}^*, \hat{\mathcal{F}}) = V(D(\hat{\mathcal{F}}) \setminus D(\mathcal{F}^*)) \cup V(D(\mathcal{F}^*) \setminus D(\hat{\mathcal{F}})), \tag{4.6}$$

and shown in Figure 4.2.

Misclassification rate. Another performance metric is the misclassification rate. This metric computes the proportion of points that are misclassified under the estimation of the Pareto set. Misclassification can occur in two forms: either a reference Pareto-optimal point is misclassified as dominated, or vice versa. Let $\mathcal{P}^*$ and $\hat{\mathcal{P}}$ represent the reference Pareto set and the estimated Pareto set obtained with the optimization algorithm, respectively. The misclassification rate, denoted by $M(\mathcal{P}^*, \hat{\mathcal{P}})$, is defined as follows:

$$M(\mathcal{P}^*, \hat{\mathcal{P}}) = \frac{1}{|\mathbb{X}|} \sum_{x \in \mathbb{X}} (\mathbb{1}_{x \in \mathcal{P}^*} - \mathbb{1}_{x \in \hat{\mathcal{P}}})^2 . \tag{4.7}$$

Correct selection. Several other performance measures exist, for example, the indicator of correct selection as used in the Bayesian ranking and selection literature (Lee et al., 2010, see, e.g.,). This metric simply monitors whether all the Pareto and non-Pareto optimal points are correctly classified.

4.2.3 . Batch evaluations

Due to the stochastic nature of the outcomes, each output Z^i is obtained as an empirical average of the outputs over several replications performed at the same point. The number of replications is commonly referred to as the batch size, denoted by k in the following. The advantage of dealing with batches of evaluations is that the GP posteriors can be updated at the same computational complexity than that of a single point (see, e.g., Binois et al., 2018).

The use of replications, for building surrogate models for noisy black-box functions, has been proved to be helpful in various studies in the literature (see, e.g., Horn et al., 2017, Jalali et al., 2017). This idea is used for instance by Binois et al. (2017) for developing sequential design schemes for stochastic simulation experiments and by Barracosa et al. (2021) for extending the Pareto Active Learning (PAL) algorithm for stochastic simulators.

In the literature of ranking and selection, Gonzalez et al. (2022) proposed a multiobjective optimization algorithm which adaptively determines the batch size, but this type of algorithms becomes unsuitable for large grid input space problems. In the analysis of PALS, Barracosa et al. (2021) provides insight into the influence of batch size over performance metrics and recommends lower batch sizes for faster reduction in metric value during initial iterations.

Given a fixed budget setup for total evaluations, the choice of batch size determines the trade-off between exploration and replication in sequential design of computer experiments (see, e.g., Binois et al., 2017).

4.3 . Maximal uncertainty sampling

4.3.1 . Maximal uncertainty sampling for function approximation

In the literature of Bayesian sequential design and analysis of computer experiments, we can distinguish two main families of approaches for constructing a sampling criterion: maximal uncertainty sampling (MUS) and stepwise uncertainty reduction (SUR).

In the following, our objective is to use the principle of MUS to design an sampling criterion for Bayesian stochastic multi-objective optimization. To make the presentation more didactic, we start with the problem of function approximation and that of the estimation of the probability of an excursion set, which are classically addressed in the literature by MUS-type principles.

The problem of function approximation using kriging-based sequential search strategies is well studied in the literature (see, e.g., Currin et al., 1991, Santner et al., 2003, Vazquez and Bect, 2011, Welch et al., 1992). In this context, the principle of MUS is illustrated Figure 4.3, where the objective is to build a surrogate model for the function represented by blue dotted lines. The red points represent the noise-free evaluations made on the true function, while a Gaussian process model yields the posterior mean μ_n, depicted by the solid blue line. Uncertainty is gauged in terms of the posterior standard deviation σ_n, as indicated by the light blue shaded 95% confidence interval around the posterior mean. The subsequent evaluation point to be sampled, indicated by the red vertical line, is chosen where the posterior variance σ_n^2 is maximal, as demonstrated in the second subfigure.

The MUS-type approach is also classical in the literature of reliability theory for the estimation of a probability of failure of a system in a Bayesian framework. Given a probability distribution $P_{\mathbb{X}}$ on the input space $\mathbb{X}$, which accounts for the uncertainty in the system input factors, and a threshold $u \in \mathbb{R}$ corresponding to a critical value, the problem is about the estimation of the probability α that f exceeds u:

$$\alpha = P_{\mathbb{X}}\{x \in \mathbb{X} \,:\, f(x) > u\}\,. \tag{4.8}$$

In addressing this problem, a natural and intuitive method for measuring uncertainty is the probability of misclassification (see, e.g., Bryan et al., 2005). Given the data I_n, let $p_n(x)$ denote the posterior probability that the point x is above the given threshold u. This can be computed using the posterior mean μ_n and posterior variance σ_n^2, of the Gaussian process model ξ_n given the data

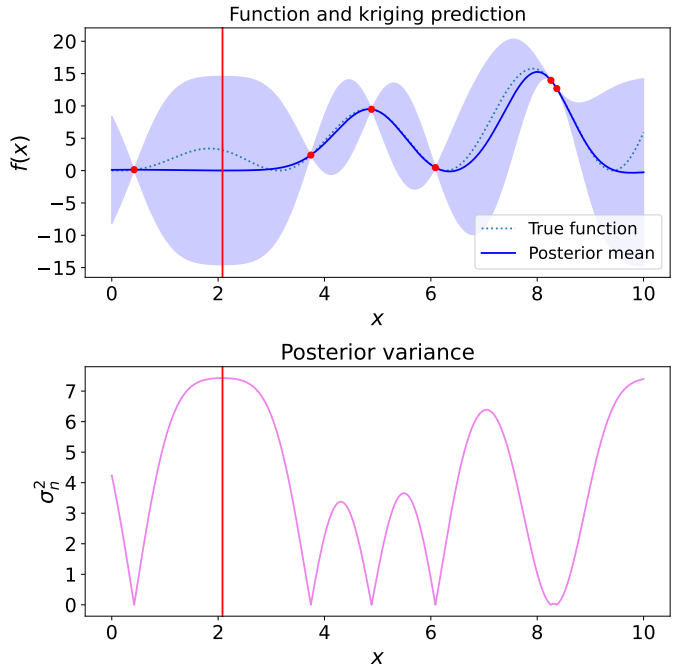


Figure 4.3: Function approximation using MUS

I_n,

$$p_n(x) = \Phi\left(\frac{\mu_n(x) - u}{\sigma_n(x)}\right), \tag{4.9}$$

where Φ is the cumulative distribution function of the normal distribution. A point $x \in \mathbb{X}$ can be predicted to be above the threshold if $p_n(x) > 0.5$, thereby deriving the sampling criterion $J_n(x) = \min\left(p_n(x), 1 - p_n(x)\right)$, which represents the probability of misclassification. By maximizing this criterion, we aim to identify the next sampling point $X^{(n+1)}$ that exhibits a probability $p_n(X^{(n+1)})$ as close as possible to $1/2$.

Figure 4.4 shows a MUS strategy based on probability of misclassification in the context of estimating α, based on the same function example as in Figure 4.3. The figure on the top shows the threshold u in red dotted lines. The sampling criterion based on the probability of misclassification, as shown in the second subfigure, is maximized to select the next evaluation point.

Remark 6. In addition to the criterion based on the probability of misclassification, there are several criteria that lead to equivalent choices of sampling points, (see, e.g., Abdelmalek-Lomenech et al., 2022) including the variance or the entropy of the indicator function for correct classification (Bryan et al., 2005), as demonstrated in Figure 4.5, or the "learning function U" as proposed by Echard et al. (2011).

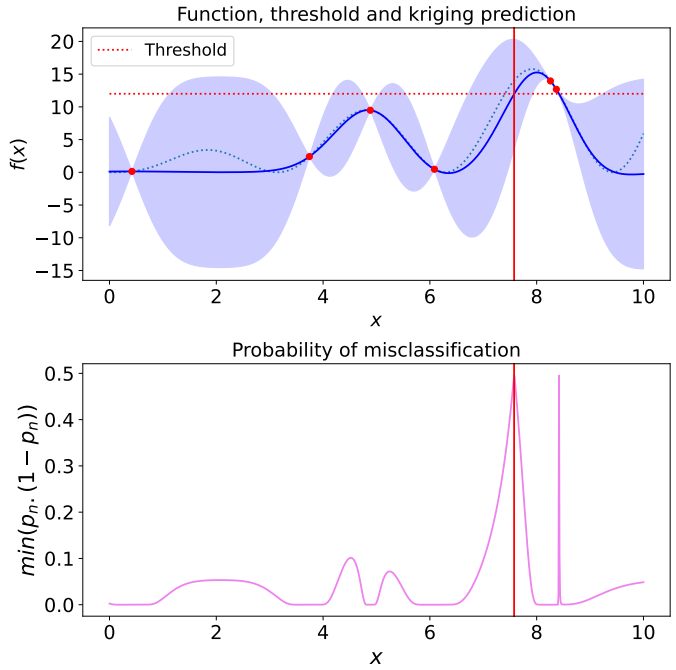


Figure 4.4: Estimation of probability of failure using MUS

4.3.2 . A first approach

In the context of BSMOO, we study the performance of MUS using sampling criteria from the literature on Gaussian processes for reliability.

As a first approach, we consider a sampling criterion given by the posterior variance of the indicator $\mathbb{1}_{x\in\mathcal{P}}$, that may be written as

$$\mathrm{var}_n\left(\mathbb{1}_{x\in\mathcal{P}}\right) = p_n(x)(1-p_n(x)), \tag{4.10}$$

where $p_n(x)$ denotes the posterior probability of $x \in \mathbb{X}$ belonging to the Pareto set $\mathcal{P}$.

To the best of our knowledge, there is no closed form for $p_n(x)$. Thus, p_n must be estimated using conditional simulations (see, e.g., Binois et al., 2015, Villemonteix et al., 2007) drawn from the Gaussian process ξ_n conditional on all the observations up to the n-th step I_n.

This sampling criterion is tested against a naive random search method, which samples the next evaluation point randomly from $\mathbb{X}$. The two methods are tested on a series of bi-objective test problems taken from Barracosa et al. (2021), and for each case, we compute the two measures of performance metrics, namely, the volume of symmetric difference and the misclassification rate, as shown in figure 4.6, for just one test case. The experiments used an initial design of size 20, an input space of dimension 21×21, and a budget of 250×200 evaluations with a batch size of $k = 200$.

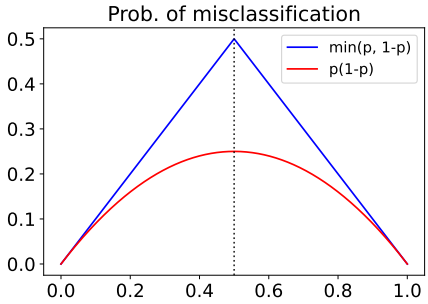
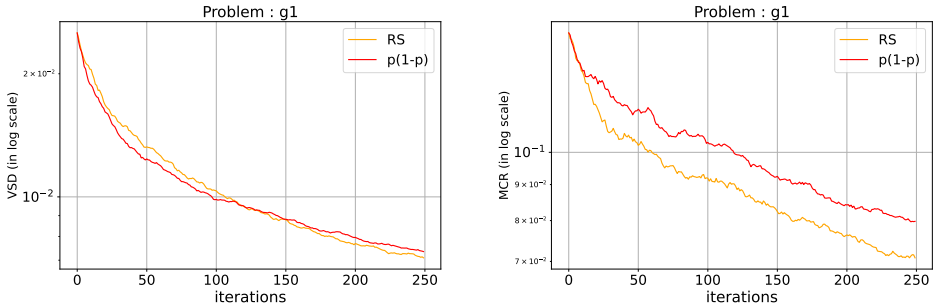


Figure 4.5: Equivalent sampling criteria for the probability of misclassification.



(a) Volume of symm. difference

(b) Misclassification rate

Figure 4.6: Performance metrics averaged over 200 independent repetitions of the experiment, on the $g1$ function from Barracosa et al. (2021), for the $p_n \cdot (1 - p_n)$ (in red) based MUS method compared to a naive Random Search (RS) (in yellow) method.

Surprisingly, the volume of symmetric difference metric shows almost no significant improvement of the experimented strategy over random search. Moreover, despite being constructed based on the probability of misclassification, the method performed worse than the random search method on the estimation of the Pareto set, as depicted by the misclassification metric. This unexpected performance of the probability of misclassification-based MUS criterion leads us to investigate the reason for these results.

Figure 4.7 displays one of the test problems named $g1$ from Barracosa et al. (2021), elucidating the typical behavior of this type of MUS criterion. The figure on the left presents the test problem $g1$ in the objective space, and the one on the right zooms in on the region indicated by the red rectangle, illustrating the relative positions of the points near the Pareto front, estimated using posterior means.

In addition to the estimated positions of the points, the figure shows the probabilities p_n of belonging to the Pareto front.

The blue points have a high probability p_n, indicating that they are likely

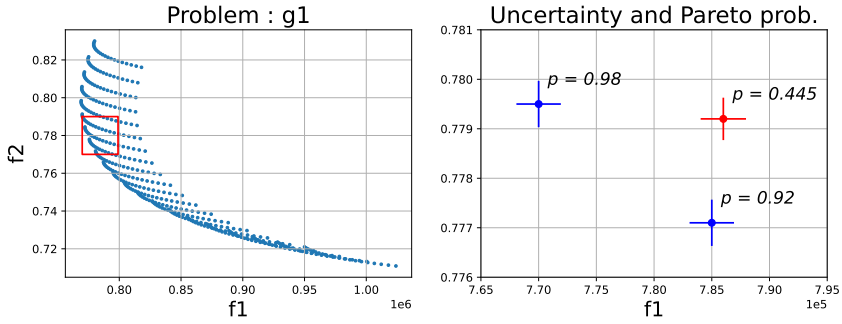


Figure 4.7: Test case showing typical positioning of the points leading to poor performance of the uncertainty measures. The left figure shows test function with the true values in the objective space, and the right figure focuses on some particular points inside the area highlighted using the red box. The right figure plots the posterior mean of the corresponding points as obtained with the GP model along with estimated probabilities $p_n(x)$ of being Pareto optimal and the associated uncertainty in terms of posterior variance, shown using horizontal and vertical lines across both axes.

to be points on the Pareto front. On the other hand, the red point has a probability p_n close to $1/2$. The MUS criterion $p_n(x)(1-p_n(x))$ is high at this point, close to $1/4$, and this red point may be selected for evaluation. On the other hand, the sampling criterion at the blue points will be small.

By sampling at the red point, the posterior variance of the red point will decrease, but that of the blue points will remain high. Because of the proximity of the blue points and the fact that their variance will remain high, since they will not be sampled, the red point will remain indeterminate as to whether it belongs to the Pareto front. Indeed, if the posterior position of the bottom right blue point were, for example, slightly further to the right, the red point would become non-dominated.

As a consequence, the algorithm will repetitively sample the red point.

Thus, the selection criterion, although it targets misclassification, will not be able to reduce it. To reduce misclassification, it is also necessary to sample the blue points, i.e., points that are already deemed as well classified.

The important conclusion here is that not all MUS criteria lead to consistent algorithms.

4.3.3 . Weighted Mean Squared Error

In this subsection, we focus on a different measure of uncertainty based on a weighted mean squared error (w-MSE) of prediction, within the same framework of MUS.

Given observations $I_n = \{(X^1, Z^1), \ldots, (X^n, Z^n)\}$, the next point $X^{n+1} \in \mathbb{X}$ to be sampled is determined by maximizing the w-MSE criterion, which is

computed with the updated Gaussian process ξ_n posteriors,

$$X^{n+1} = \text{argmax}_{x \in \mathbb{X}} \Big(w_n(x) \cdot \sum_{j=1}^{q} \frac{\sigma^2_{j,n}(x)}{s^2_{j,n}} \Big), \tag{4.11}$$

where, for each point $x \in \mathbb{X}$ and the j-th objective, $j = 1, 2, \ldots, q$, $\sigma^2_{j,n}(x)$ represents the posterior variances and $s^2_{j,n}$ are normalizing constants ensuring equal scaling for all objectives. The weights $w_n(x)$ are introduced to prioritize the "potentially Pareto optimal" points and place more emphasis on them than on the points distant from the Pareto front.

PAL(S) algorithms. Two instances of w-MSE sampling criteria for BSMOO is the Pareto Active Learning (PAL) algorithm proposed by Zuluaga et al. (2013) and its extension PALS for stochastic simulators (Barracosa et al., 2021). These two algorithms use binary (0–1) weights.

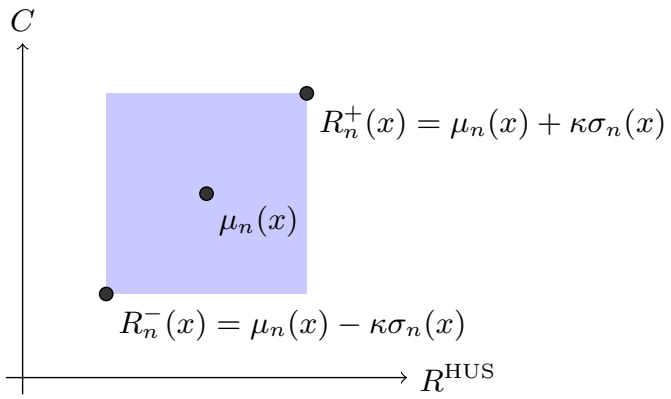


Figure 4.8: PALS rectangle with optimistic $R_n^-(x)$ and pessimistic $R_n^+(x)$ outcomes for input x.

During the n-th iteration, the PAL(S) algorithm constructs confidence rectangles around each point $x \in \mathbb{X}$ in the objective space, as depicted in Figure 4.8, using the posterior mean $\mu_n(x) = (\mu_{j,n}(x))_{1 \le j \le q}$, variances $\sigma^2_n(x) = (\sigma^2_{j,n}(x))_{1 \le j \le q}$, and a confidence parameter κ. Each rectangle is associated with both an optimistic $R_n^-(x)$ and a pessimistic $R_n^+(x)$ outcome based on the posterior uncertainty of x (see definition on Figure 4.8). The algorithm assigns the weights $w_n(x) = \mathbb{1}_{x \in \mathbb{X} \setminus N_n}$, where the set N_n corresponds to the points estimated as dominated:

$$N_n = \{ x \in \mathbb{X} \mid \exists x' \in \mathbb{X} \setminus \{x\}, R_n^+(x') \prec R_n^-(x) \}. \tag{4.12}$$

In other words, a point $x \in \mathbb{X}$ will be included in the set N_n if the corresponding optimistic outcome is dominated by the pessimistic outcome of any other point.

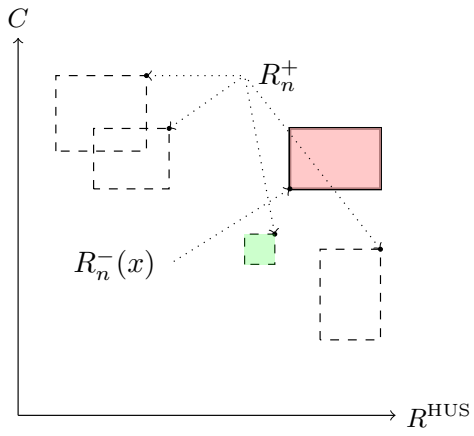


Figure 4.9: Figure illustrating the PALS classification rule, with the red box corresponding to a member of the set N_n and the green box corresponding to a potentially Pareto optimal point.

For instance, in Figure 4.9, the optimistic outcome of the red box is dominated by the pessimistic outcome of the green box, indicating that the point corresponding to the red box is a member of the set N_n. The rationale behind the choice of weights in PAL(S) is to sample more densely in regions closer to the Pareto front, rather than in regions with dominated points.

Generalization. We introduce several alternative weight functions, as presented in Table 4.1.

Table 4.1: Different weight functions for w-MSE methods

PALS	w-MSE-I	w-MSE-II	w-MSE-III
$\mathbb{1}_{x\in\mathbb{X}\setminus N_n}$	$\mathbb{1}_{x\in\hat{\mathcal{P}}_n^\alpha}$	$p_n(x)$	$p_n(x)\cdot(1-p_n(x))$

The proposed weight functions are derived from $p_n(x)$, for $x \in \mathbb{X}$, the probability of being Pareto optimal, which is estimated using conditional simulations from the Gaussian process model at the n-th iteration.

The first method, w-MSE-I, is grounded on the computation of the set $\hat{\mathcal{P}}_n^\alpha$ defined as

$$\hat{\mathcal{P}}_n^\alpha = \{x \mid p_n(x) > \alpha \cdot \min_{x\in\hat{\mathcal{P}}_n} p_n(x)\}\,, \tag{4.13}$$

where $\hat{\mathcal{P}}_n$ is the estimated Pareto set using the posterior means of ξ_n. This method selects a fraction of points in $\mathbb{X}$ whose probability of being Pareto optimal high enough relatively to the minimum value of this probability.

The other two weight functions, w-MSE-II and w-MSE-III, correspond to the probability of being Pareto optimal and the probability of misclassification, respectively.

4.3.4 . Numerical experiments

In this section, we present a numerical benchmark of the performance of the proposed w-MSE methods against PALS and random search.

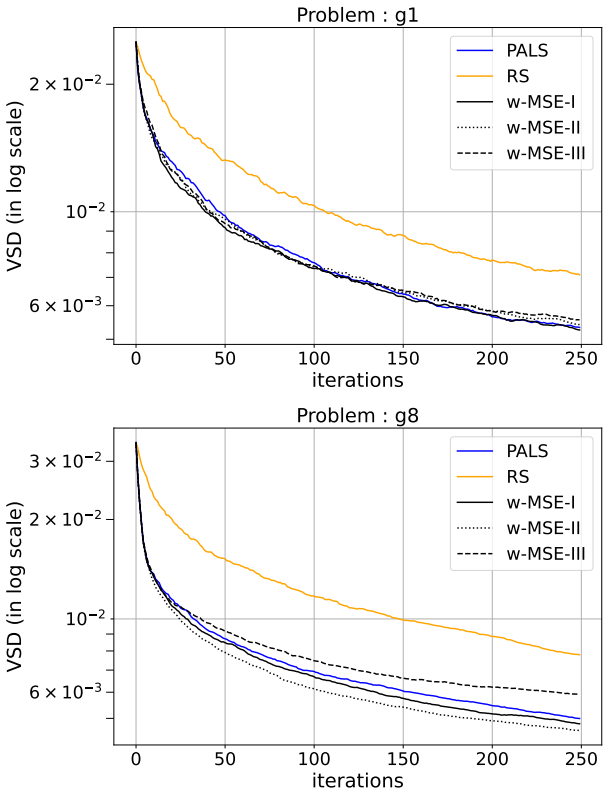


Figure 4.10: Average volume of symmetric difference (VSD) metric for PALS (in blue), random search (in yellow), and three tested w-MSE methods (in black) on test problems $g1$ and $g8$.

The algorithms are tested across a finite input space of 21×21 dimensions, involving nine bi-objective and bi-dimensional test problems adapted from Barracosa et al. (2021). These experiments were conducted under the influence of additive homoscedastic Gaussian noise and for each test problem and optimization method, a total of 200 optimization runs were executed. To assess the performance on the estimation of the Pareto front, we compute the volume of symmetric difference metric.

The GP models used for the two objectives were independent with a Matérn $5/2$ kernel and an unknown constant mean function and plug-in estimates are

used for the parameters of the covariance function. The confidence interval to construct the rectangles of the PALS algorithm was kept at 50%, as recommended in Barracosa et al. (2021).

An initial design was obtained by performing 200 evaluations at 20 randomly selected points and in addition to that a budget of 250×200 is allowed, which translates into a batch size of 200 evaluations for 250 iterations. The Pareto front at the end of the iterative procedure is estimated using the posterior mean of the GP model.

Figure 4.10 shows the results of the numerical experiments only for the test functions $g1$ and $g8$, and the remaining results are available in Appendix C.

In terms of performance with respect to the VSD metric, the proposed w-MSE methods did not show significant and consistent improvement over the PALS algorithm, considering all the test problems.

This motivates the exploration of the other branch of constructing sampling criterions, as discussed in the next section.

4.4 . Weighted integrated mean squared error sampling criterion

In this section, we delve into a family of sampling criteria referred to as stepwise uncertainty reduction (SUR). The concept of SUR was introduced for sequential design of numerical experiments by Vazquez and Bect (2009), Vazquez and Martinez (2006), Villemonteix et al. (2007) and stems from the idea of stepwise entropy reduction applied to shape recognition (Geman and Jedynak, 1996), as well as from active learning Cohn et al. (1996), MacKay (1992). Bect et al. (2019) provides a review of SUR methods.

The idea is to construct a statistics H_n measuring the uncertainty about the true Pareto front $\mathcal{F}$, based on the estimated Pareto front $\hat{\mathcal{F}}_n$ given all the observations up to the n-th step.

Given the data $I_n = \{(X^1, Z^1), \ldots, (X^n, Z^n)\}$ the sampling criterion of a SUR strategy may be written as

$$J_n(x) = \mathbb{E}_n(H_{n+1} \mid X^{n+1} = x), \tag{4.14}$$

where $\mathbb{E}_n(\cdot \mid X^{n+1} = x)$ denotes the conditional expectation given I_n, assuming $X^{n+1} = x$.

This sampling criterion should be minimized to select a new evaluation point:

$$X^{n+1} = \operatorname{argmin}_{x \in \mathbb{X}} J_n(x)\,. \tag{4.15}$$

Depending on the uncertainty measure H_n, the computation of the conditional expectation $\mathbb{E}_n$ is not always analytically tractable. In this case, methods based on conditional simulations may be used.

In this work, we propose a weighted integrated mean squared error (w-IMSE) based uncertainty measure,

$$H_n = \sum_{x_i \in \mathbb{X}} w_n(x_i) \sum_{j=1}^{q} \sigma^2_{j,n}(x_i).$$

This yields the following SUR sampling criterion:

$$J_n(x) = \sum_{x_i \in \mathbb{X}} w_n(x_i) \sum_{j=1}^{q} \frac{\sigma^2_{j,n+1}(x_i \mid x)}{s^2_{j,n}}\,, \tag{4.16}$$

where $\sigma^2_{j,n+1}(\cdot \mid x)$ is the predictive posterior variance at step $n+1$ when a new observation at $X^{n+1} = x$ is added, and $s^2_{j,n}$ are normalizing constants ensuring equal scaling for all objectives. The weights w_ns are computed separately and used in a plug-in approach to construct the sampling criterion. Note that since $\mathbb{X}$ is discrete, this integral criterion boils down to a sum over the points of $\mathbb{X}$.

Using the properties of Gaussian processes, it can be shown that (Emery, 2009) the new predictive distribution can be expressed simply as a function of the predictive posterior variance $\sigma^2_{j,n}$ and covariance $\mathrm{cov}_{j,n}(x_i, x)$ at step n. Moreover, the variance term $\sigma^2_{j,n+1}(x_i \mid x)$ does not depend upon the new observation value:

$$\sigma^2_{j,n+1}(x_i \mid x) = \sigma^2_{j,n}(x_i) - \frac{\mathrm{cov}^2_{j,n}(x_i, x)}{\sigma^2_{j,n}(x)}. \tag{4.17}$$

4.4.1 . Proposed method

The weights for the w-IMSE SUR criterion could be chosen from Table 4.1, with an emphasis on potentially Pareto optimal points. Nevertheless, empirical studies not included in this manuscript suggest that transitioning from w-MSE to w-IMSE, using these particular weights, does not lead to significant enhancements in performance.

Instead, we propose a new weight function that focuses on the estimation of the Pareto front in terms of errors measured by the volume of symmetric difference.

Like in the PALS approach, we use the classification of the dominated class of points N_n to assign non-zero weights to selected points.

Let $\mathcal{P}_n$ denote the estimated Pareto set at the n-th iteration, and let the corresponding estimate of the Pareto front $\mathcal{F}_n$ be defined as

$$\hat{\mathcal{F}}_n = \{\mu_n(x) \mid x \in \mathcal{P}_n\}\,, \tag{4.18}$$

where $\mu_n(x)$ denotes the posterior means for $x \in \mathbb{X}$ from the GP model ξ_n.

Next, we define a pessimistic and an optimistic Pareto front, denoted by $\mathcal{F}_n^+$ and $\mathcal{F}_n^-$ respectively, such that

$$\mathcal{F}_n^+ = \{R_n^+(x) \mid x \in \mathcal{P}_n\}\,, \tag{4.19}$$

$$\mathcal{F}_n^- = \{R_n^-(x) \mid x \in \mathcal{P}_n\}\,, \tag{4.20}$$

where $R_n^+(x) = \mu_n(x) + \kappa\sigma_n(x)$ correspond to pessimistic values of the objectives and $R_n^-(x) = \mu_n(x) - \kappa\sigma_n(x)$ correspond to optimistic values, as in the PAL(S) algorithms. The pessimistic and optimistic fronts are depicted in the left plot of Figure 4.11.

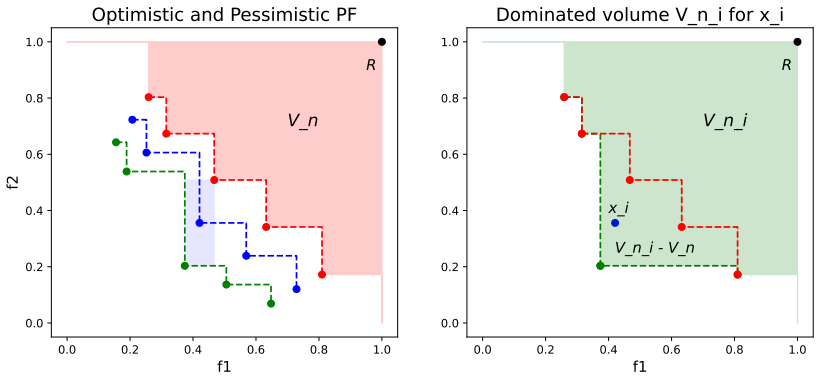


Figure 4.11: Left plot: Optimistic $\mathcal{F}_n^-$ (in green), pessimistic $\mathcal{F}_n^+$ (in red), and the estimated $\mathcal{F}_n$ Pareto front (in blue). Right plot: Dominated volume V_n by $\mathcal{F}_n^+$ and $V_{n,i}$ dominated by $\mathcal{F}_n^*$.

Given a reference point R, we introduce an operator D to denote the volume of the region dominated by a Pareto front. Then, we denote by V_n the volume of the region dominated by $\mathcal{F}_n^+$. : $V_n = D(\mathcal{F}_n^+)$.

For each point $x_i \in \mathbb{X}$, the weight within the w-IMSE criterion is calculated based on a potential contribution to the increase of V_n, more specifically,

$$w_n(x_i) = (D(\mathcal{F}_{n,i}^*) - V_n)/V_n\,, \tag{4.21}$$

where $\mathcal{F}_{n,i}^*$ is a Pareto front obtained by replacing the pessimistic value $R_n^+(x_i)$ by its optimistic counterpart $R_n^-(x_i)$ in $\mathcal{F}_n^+$, which may be formally defined as follows:

$$U_i = \{\mathcal{F}_n^+ \setminus R_n^+(x_i)\} \cup R_n^-(x_i), \tag{4.22}$$

$$\mathcal{F}_{n,i}^* = \{y \in U_i : \nexists y' \in U_i \text{ s.t. } y' \prec y\}\,. \tag{4.23}$$

The motivation behind this approach is similar to PAL(S), which tries to reduce the uncertainty around the Pareto front. In two dimensions this uncertainty can be visualized as the confidence interval generated by the optimistic ($\mathcal{F}_n^-$) and pessimistic ($\mathcal{F}_n^+$) Pareto fronts. The proposed algorithm gives

Algorithm 2 Construction of w_n at step n

Estimate Pareto set $\widehat{\mathcal{P}_n}$ based on GP posterior mean
$\mathcal{F}_n^+ \leftarrow \{R_n^+(x) | x \in \widehat{\mathcal{P}_n}\}$ ▷ (Pessimistic Pareto Front)
$V_n \leftarrow D(\mathcal{F}_n^+)$ ▷ (Reference dominated volume)
for $x_i \in \mathbb{X}$ **do**
  if $x_i \in N_n$ **then**
    $w_n(x_i) \leftarrow 0$
  else
    $\mathcal{F}_n^* \leftarrow$ Pareto front of $\{\mathcal{F}_n^+ \cup R_n^-(x_i)\}$
    $V_{n,i} \leftarrow D(\mathcal{F}_{n,i}^*)$ ▷ (Dominated volume)
    $w_n(x_i) \leftarrow (V_{n,i} - V_n)/V_n$ ▷ (normalized weights)
  end if
end for
Return w_n

maximum weight to the point which contributes the most in this uncertainty in terms of dominated area.

4.4.2 . Numerical experiments

Following the same framework of numerical experiments as described in Section 4.3.4, the performance of the proposed w-IMSE algorithm is tested against PALS and the random search method. Figure 4.12, shows the VSD metric for the methods, corresponding to all the test problems.

The proposed algorithm, shown in red, shows consistent improvement over PALS for all the considered test problems, with respect to the VSD metric. However the proposed algorithm does not show any improvement over the PALS algorithm for the estimation of the Pareto set, as monitored by the misclassification rate in Appendix C. The proposed weights in Algorithm 2 were also tested in the w-MSE framework, which showed slightly improved performance with respect to the VSD metric, over the w-IMSE variant. But at the same time the w-MSE algorithm performed very poorly for the estimation of the Pareto set as monitored by the misclassification rate, demonstrated by the results in Appendix C.

4.5 . Application to QMRA simulator

4.5.1 . Problem formulation

This study of multiobjective optimization algorithms is motivated by the problem of optimizing intervention parameters in raw milk cheese production. This section demonstrates a simple example of the optimization problem based on the single pathogen QMRA model for MPS-STEC. The aim is to

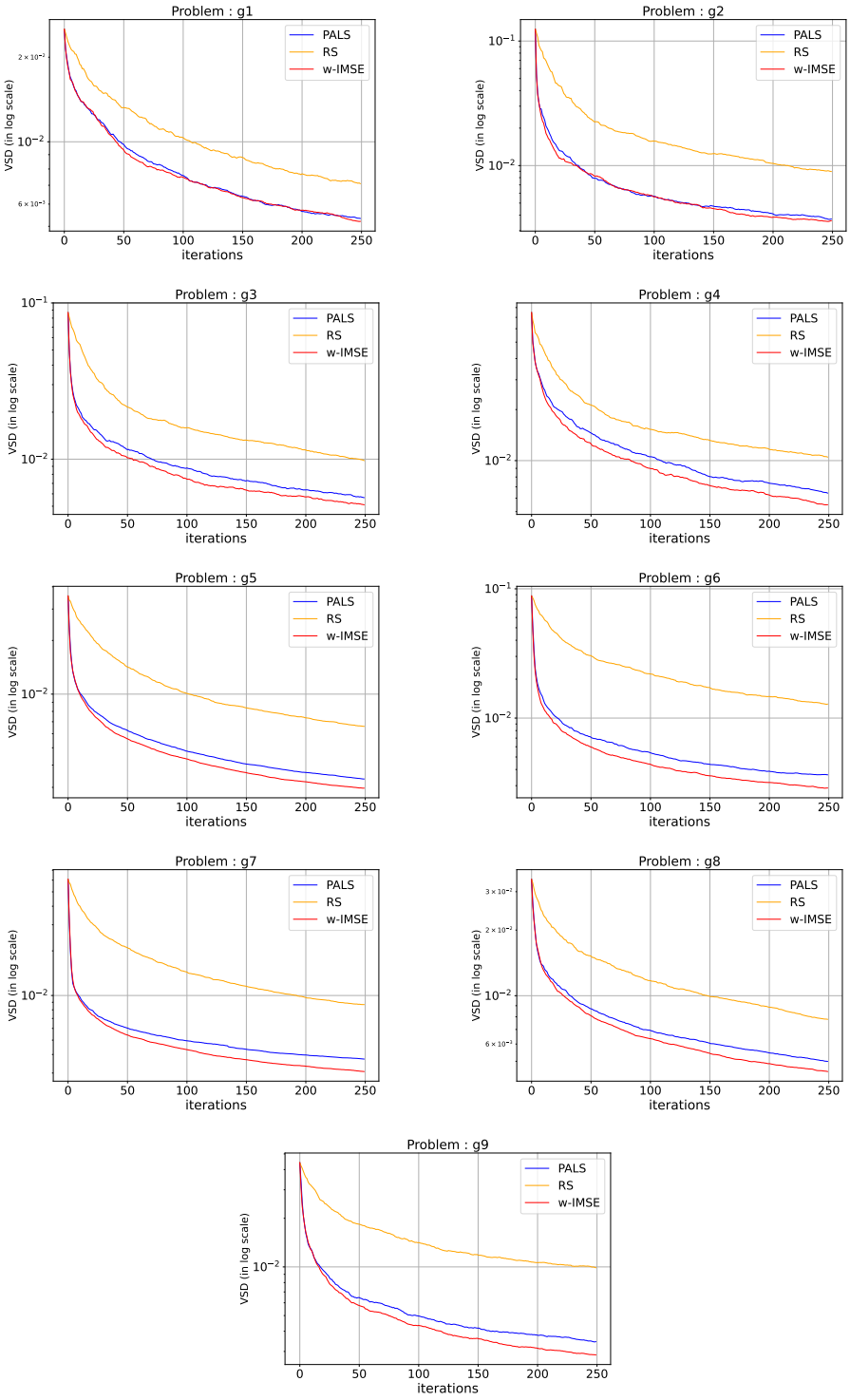


Figure 4.12: Average VSD metric on test problems from Barracosa et al. (2021).

minimize two conflicting objectives or quantities of interest, namely, the relative risk of HUS and the cost of intervention processes. Given a specific computational budget, these quantities of interest can be computed to a certain degree of accuracy using the computationally expensive (see Section 2.3.2.1) and stochastic QMRA model. Among all the inputs of the MPS-STEC QMRA model, only four input parameters corresponding to the intervention steps are selected to be optimized and these parameters are collectively denoted by $\theta^{\text{opt}} = \{l^{\text{sorting}}, p^{\text{milk}}, n^{\text{sample}}, p^{\text{cheese}}\}$. l^{sorting} denotes the maximum threshold of pathogen concentration for milk testing, $p^{\text{milk}} = \frac{1}{f^{\text{sorting}}}$ is the proportion of farm milk tested, n^{sample} denotes the number of cheese sample tested and p^{cheese} is the proportion of cheese batches tested. All the other input parameters of the MPS-STEC QMRA model are kept fixed for this experiment. The input space considered for the optimization problem is discrete and finite, denoted by $\mathbb{X} = \theta_1^{\text{opt}} \times \theta_2^{\text{opt}} \times \theta_3^{\text{opt}} \times \theta_4^{\text{opt}}$, consisting of all the possible 1500 combinations of values of the parameters, as demonstrated in the Table 4.2.

Table 4.2: Inputs of the QMRA simulator to be optimized θ^{opt}.

Symbol	Values
l^{sorting}	$\theta_1^{\text{opt}} = [10, 20, 30, 50, 100, 200]$
p^{milk}	$\theta_2^{\text{opt}} = [0.1, 0.2, \ldots, 1]$
n^{sample}	$\theta_3^{\text{opt}} = [3, 5, 8, 10, 12]$
p^{cheese}	$\theta_4^{\text{opt}} = [0.1, 0.2, \ldots, 0.5]$

For a particular input $x \in \mathbb{X}$, the QMRA model produces three outputs corresponding to the production of a particular batch of cheese, namely, 1) $Z_1(x) \coloneqq R^{\text{batch}}_{\text{MPS-STEC}}(\Xi_{\text{MPS-STEC}})$, the probability of getting HUS from consuming a portion of 25g of cheese from a particular batch, characterized by the stochastic internal variables $\Xi_{\text{MPS-STEC}}$, 2) $Z_2(x) \coloneqq P^{\text{batch}}(\Xi_{\text{MPS-STEC}})$, the probability of detecting MPS-STEC contamination in at least one of the n^{sample} cheese samples of 25g, sampled from that batch of cheese characterized by $\Xi_{\text{MPS-STEC}}$ and 3) $Z_3(x) \coloneqq M^{\text{batch}}(x)$, the quantity of milk (in Liters) lost due to milk testing at the farm level. The QMRA model is a stochastic simulator, or in other words, for a particular input x it simulates a set of stochastic internal variables $\Xi_{\text{MPS-STEC}}$, that eventually affects the three outputs. As a result all the three outputs are random variables, that is, for a fixed x the output values are observed with noise.

As mentioned before, this section only concerns about a simple version of the multiobjective optimization problem in the context of the ArtiSaneFood project and this involves minimization of two objectives related to MPS-STEC only. However depending on the choices of the decision maker or cheese

industrials, a more complicated optimization problem can be formulized using the multipathogen QMRA model by considering the combined effects of all the pathogens using DALYs (see Section 2.2.6.2). In the particular example under consideration, the two quantities of interest that are subjected to minimization, are the relative average risk of HUS $R(x)$ and the total cost of all the intervention steps $C(x)$,

$$\begin{aligned} R(x) &= \frac{\mathbb{E}[Z_1(x)(1 - Z_2(x)p^{\text{cheese}})]}{(1 - \mathbb{E}[Z_2(x)p^{\text{cheese}}])\, R^{\text{baseline}}_{\text{MPS}-\text{STEC}}} \\ C(x) &= \mathbb{E}[c + c_1 \cdot Z_3(x) + c_2 \cdot Z_2(x)], \end{aligned} \tag{4.24}$$

where, $c = N^{\text{farm}} C^{\text{milk}}_{\text{test}} p^{\text{milk}} + n^{\text{sample}} C^{\text{cheese}}_{\text{test}} p^{\text{test}}$, $c_2 = p^{\text{test}} C^{\text{cheese}}_{\text{loss}} N^{\text{cheese}}$ and $c_1 = C^{\text{milk}}_{\text{loss}}$ are constants used to compute the cost values (see Section 2.9) and $R^{\text{baseline}}_{\text{MPS}-\text{STEC}} := \mathbb{E}[Z_1(x^{\text{baseline}})]$ is the average risk of HUS in a baseline scenario signifying intervention steps, or equivalently with input $x^{\text{baseline}} = \{l^{\text{sorting}} = \infty,\, p^{\text{milk}} = 0,\, n^{\text{sample}} = 0,\, p^{\text{cheese}} = 0\}$.

4.5.2 . Modified PALS algorithm

It is to be noted that among the two quantities of interest as defined in (4.24), unlike $C(x)$, the other quantity $R(x)$ can not be expressed as an expectation of some function of the simulator outputs. In the multiobjective stochastic optimization framework presented in this chapter, Gaussian process surrogate models are constructed for the quantities of interest that are subjected to optimization. In such a framework the quantities of interest are the expected values of the simulator outputs, which allows to use batch evaluations (see Section 4.2.3), to address the presence of noise in the observations. The optimization problem of the QMRA model, as presented in this section, exhibits a different scenario, due to definitions of the quantities of interest in (4.24). To solve this applicability issue a modified version of the PALS algorithm (Barracosa et al., 2021) is proposed by Basak et al. (2022a). Here, we particularly work with the PALS algorithms because it has a w-MSE based sampling criterion, which is easily adaptable with the proposed modifications and also does not include any computationally expensive steps. Similar modifications with the proposed w-IMSE based algorithm is left as a future perspective.

In the algorithm proposed by Basak et al. (2022a), the Gaussian process surrogates are built on the simulator outputs, using batch evaluations. The GP posterior mean and variances corresponding to different simulator outputs thus obtained, are used to estimate the quantiles of the quantities of interest. At a particular iteration n, the construction of the dominated set N_n remains the same as the original PALS algorithm, as described in (4.12), except for the construction of the confidence rectangles. For the j-th, $j = 1, 2, \ldots, q$ quantity of interest, both the optimistic $R^{-}_{n,j}(x)$ and pessimistic $R^{+}_{n,j}(x)$ vertices of

the confidence rectangle is computed using quantiles of a certain coverage probability α,

$$\begin{aligned} R^{-}_{n,j}(x) &= Q^{\alpha/2}_{n,j}(x), \\ R^{+}_{n,j}(x) &= Q^{1-\alpha/2}_{n,j}(x). \end{aligned} \tag{4.25}$$

For the n-th iteration at input $x \in \mathbb{X}$, $Q^{q}_{n,j}(x)$ denotes the quantile of order q of the j-th quantity of interest. To use the same confidence level as in the original PALS rectangles, the confidence parameter α can be chosen with respect to the confidence parameter κ, in the PALS algorithm, using the equivalence relation $\alpha = 1 - 2(1 - \Phi(\kappa))$. The next sampling point selected based on the MUS sampling criterion for this modified algorithm is therefore defined as,

$$X_{n+1} = \text{argmax}_{x \in \mathbb{X} \setminus \mathbb{N}} \| R^{-}_{n}(x) - R^{+}_{n}(x) \|^2. \tag{4.26}$$

4.5.3 . Numerical experiment

The modified PALS algorithm (Basak et al., 2022a), as described above, was used to estimate the Pareto optimal values of the parameters θ^{opt}, minimizing the two quantities of interests produced by the QMRA model of MPS-STEC. For comparison of results we use a reference Pareto front, estimated with simple Monte Carlo method, with a batch size 5000 at each of the 1500 input points of $\mathbb{X}$. The Monte Carlo batch size for the convergence of the simulator outputs, was determined based on the recommendations in FDA/CFSAN et al. (2021), which ensures the variance of the running mean is less than a specified threshold (1%, default). Figure 4.13 plots the cost and the relative risk, estimated with the Monte Carlo method. The reference Pareto front as shown in red, is considered as the true Pareto front for comparing the estimates obtained with the experimental results in this section.

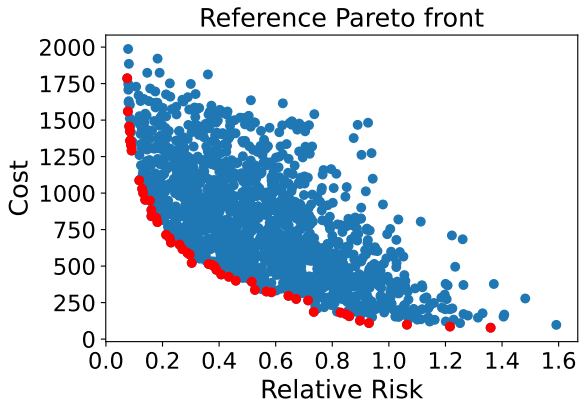
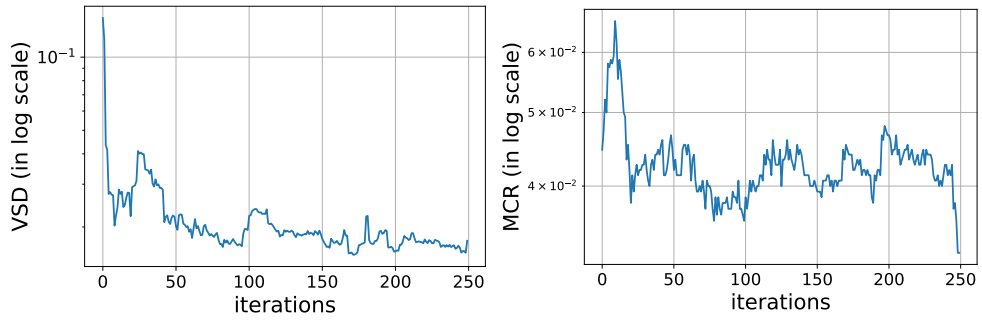


Figure 4.13: Pareto front estimated with simple Monte Carlo.

An initial design of size of 50×200 was used and the algorithm was run for 250 iterations, with a fixed batch size of 200 for each evaluation. For each

of the GP models, the noise variance was assumed to be homoscedastic and known. They were estimated offline using the 5000 Monte Carlo simulations for each $x \in \mathbb{X}$, and then averaged over all the points in $\mathbb{X}$. The GP models used the same framework as described in Section 4.3.4, with plug-in covariance hyperparameters estimated using the whole reference data. The VSD and MCR metric were computed for all the iterations of the algorithm, as shown in Figure 4.14. The algorithm uses 300×200 simulations, compared to the Monte Carlo approach, that uses 1500×5000 simulations, for estimating the Pareto optimal solutions. This demonstrates substantial reductions in computational time while incurring a marginal 4% misclassification rate for the Pareto set.



(a) Volume of symm. difference

(b) Misclassification rate

Figure 4.14: Performance metrics for a single run of the modified PALS algorithm on the single pathogen QMRA simulator.

The estimated Pareto front as obtained with the modified PALS algorithm, is shown by the red points in Figure 4.15. As discussed earlier, in multiobjective optimization problems with conflicting objectives, the solution set consists a set of optimal choices rather than a single optimal solution. Once the corresponding Pareto set, that is the optimal choices of the input parameters are obtained, the decision makers are allowed to choose among them, balancing the trade-off between the two quantities of interest.

4.6 . Discussions and perspectives

In the context of the ArtiSaneFood project, one of the main objectives was to make recommendations to french cheese producers in order to implement efficient intervention strategies. This essentially involves the joint optimization of several conflicting objectives of the QMRA model, which are stochastic in nature and computationally expensive. This type of problems are studied in the literature of BSMOO, which was the main motivation of this chapter. In this chapter we have provided a global overview of the BSMOO literature, pro-

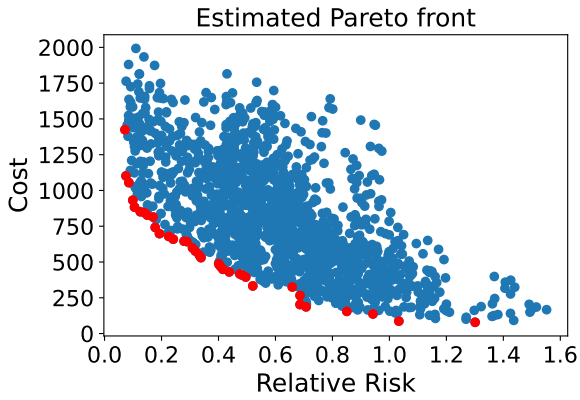


Figure 4.15: Pareto front estimated with modified PALS algorithm.

posed an extension to the PALS algorithm (Barracosa et al., 2021) adapted for our QMRA application case (Basak et al., 2022a) and proposed a new algorithm improving over the PALS algorithm on the estimation of the Pareto front. In particular we studied two main branches of constructing a sampling criterion for BSMOO algorithms, namely 1) Maximal Uncertainty Sampling (MUS) and 2) Stepwise Uncertainty Reduction (SUR). Firstly, our study shows that, uncertainty measures from the literature of reliability theory is not always adaptable in the BSMOO framework. Later we studied weighted mean squared error (w-MSE) based uncertainty measures and tested them on a numerical benchmark, against the PALS algorithm. The results concluded that, w-MSE based other sampling criterions were not significantly better than PALS. As the next contribution of this work, we proposed a new algorithm based on the weighted integrated mean squared error (w-IMSE) criterion, which performs consistently better than the PALS algorithm, for the estimation of Pareto front. This is supported by the fact that the weights in our new algorithm is based on the dominated region in the objective space, that in turns focuses on the estimation of the Pareto front.

As future perspectives of this work, there are many potential directions.

1) A first step would be to compare the proposed algorithm with alternative approaches from the literature of Bayesian Ranking and selection (see, e.g., Lee et al., 2010, Rojas Gonzalez et al., 2020). This group of algorithms is specifically tailored for discrete input spaces but are often incompatible with large input spaces. Another potential algorithm to be included the numerical benchmark, is the entropy based PESMO method proposed by Hernández-Lobato et al. (2014).

2) The numerical benchmarks presented in this study, compares algorithms with fixed hyperparameters for the covariance kernels of the GP models. This configuration was imposed to focus only on the effectiveness of the sampling criterion, on the algorithm's performance. Nonetheless, hyperpa-

rameter estimation of GP models (see, e.g., Basak et al., 2022c), remains a separate topic, which offers an interesting direction for future exploration, in the context of BSMOO algorithms.

3) The test problems considered in this study (Barracosa et al., 2021), assumes homoscedastic noise variance and it is assumed to be known for the algorithms, that are compared. This is not often the case in reality, but the noise variance can be estimated using the batches of simulation and updated at each iteration. This motivates another future aspect for extending the numerical benchmarks, by considering test problems with heteroscedastic and unknown noise variance.

4) The new w-IMSE based algorithm is proposed in a bi-objective framework, using the weights calculated, based on the area of the dominated region, on the objective space. A next possible extension of this algorithm is defining the weight function in a multiobjective setup.

5) The proposed algorithm performs better than PALS with respect to the VSD metric on the Pareto front, but concerning the estimation of the Pareto set, it does not show significant improvement on PALS. Several other weight functions, as shown in Table 4.1, were tested in the w-IMSE framework as well, but the results showed no significant improvements. A potential future direction to explore would be the analytically tractable SUR algorithms.

6) This chapter demonstrates a simple example of the QMRA optimization problem, with the single pathogen model for MPS-STEC. However the objective functions of the optimization problem and the corresponding input parameter space can vary depending on the concerned cheese producers and interests of the decision makers. In practice there can be other formulations of the optimization problem using DALYs (see Section 2.2.6.2), that include the effects of other pathogens as well. Another perspective in the context of the QMRA application includes the extension of the proposed w-IMSE based sampling criterion, when dealing with specific quantities of interest, which can not be expressed as an expectation of functions of the simulator outputs.

5 - Conclusions and Perspectives

5.1 . Contributions

This thesis focuses on the objectives of the French consortium in the European project ArtiSaneFood, which aims to support French cheese producers in implementing effective process intervention strategies.

Specifically, the research advances in three directions: 1) developing a new multipathogen quantitative risk assessment (QMRA) model for raw milk soft cheese; 2) exploring numerical integration methods for monotone bounded functions to potentially enhance the QMRA model's implementation; and 3) contributing to Bayesian optimization literature for stochastic, resource-intensive simulators.

Quantitative Risk Assessment Model. The first component of this thesis presents a novel multipathogen risk assessment model for raw milk cheese. To our knowledge, this approach is unprecedented in this context. Contrasting with many single-pathogen QMRA studies (see, e.g., Campagnollo et al., 2018, Collineau et al., 2020, Fritsch et al., 2018, Perrin et al., 2014, Sanaa et al., 2004, Strickland et al., 2023, Tenenhaus-Aziza et al., 2014), our work extends the existing QMRA model for STEC (Perrin et al., 2014) and provides an implementation in R and FSKX Basak et al. (2024). This implementation enhances the original model Perrin et al. (2014) in several ways: 1) it employs the proportion of MPS STEC directly in the farm module to calculate the number of infected cows; 2) it uses a Bayesian approach, detailed in Appendix A, for individual farm hygiene parameter estimation; 3) it provides cost estimates for preharvest and postharvest interventions; 4) it employs an efficient method for batch risk calculation using its analytical form; and 5) it updates several QMRA model input parameters, validated by dairy industry stakeholders, CNIEL and ANSES.

Chapter 2 extends the single-pathogen model Basak et al. (2024) to a multipathogen model as presented in Basak et al. (2023b). The multipathogen model incorporates existing literature and field data from CNIEL and ACTALIA, gathered under the ArtiSaneFood project. Using production parameters (θ), it estimates the risk of three food-borne illnesses: Haemolytic Uremic Syndrome (HUS) from STEC, salmonellosis from *Salmonella*, and listeriosis from *Listeria monocytogenes*. To model also estimates the burden on public health using a metric based on Disability Adjusted Life Years (DALYs) due to food-borne illnesses linked to consumption of raw milk cheese. Additionally, the model calculates the total analytical and liability costs for rejecting batches of milk/cheese as a result of intervention steps.

Implemented in the R language on the single-pathogen version Basak et al. (2024), this model aids French cheese producers in assessing food-borne illness risks from specific cheese batches. It can also facilitate scenario-based analyses or, more intriguingly, optimize relevant parameters as discussed in Chapter 4 of this thesis.

Integration of Bounded Monotone Functions. The second task, explored in Chapter 3, focuses on the integration methods for monotone bounded functions, a crucial aspect in the QMRA model implementation from the previous chapter. We specifically address integrating a function $f : [0, 1] \to [0, 1]$ with respect to a uniform random variable on the unit interval. The chapter begins with a review of existing integration methods, broadly categorized into nonsequential and sequential approaches. Firstly, we establish a lower bound for the maximum L^p error, for any $p \geq 1$, for nonsequential randomized methods. We examine two unbiased methods from this category based on control variate and stratification principles (Basak et al., 2022b). As noted by Novak (1992), nonsequential randomized methods show limited improvement, with a L^1 error lower bound of $1/8n$ compared to the deterministic trapezoidal rule's worst-case error of $1/2(n + 1)$ (see, e.g., Kiefer, 1957). Nevertheless, studying these methods is valuable, as unbiased nonsequential Monte Carlo methods serve as building blocks for the design of sequential Monte Carlo methods, which are shown (Novak, 1992) to be rate-optimal with a minimax rate of $n^{-3/2}$ for the L^1 error. In Chapter 3's second part, we explore a two-stage sequential method based on stratified sampling, as used by Novak (1992) to demonstrate the rate optimality of sequential randomized methods. Our contribution includes formalizing Novak's method into an integration algorithm suitable for our QMRA model, particularly for step-like functions where most variance is concentrated in a small domain interval. Specifically, we propose three enhancements to Novak's method: 1) a tighter upper bound for the L^1 error, reducing it by a factor of $\sqrt{2}$; 2) a novel budget allocation scheme for more efficient resource distribution over the function's domain, considering its flat regions; and 3) an optimal rule for selecting the initial number of equispaced strata for worst-case scenarios. We test these algorithms and improvements against the simple Monte Carlo method, the trapezoidal rule, and Novak's original method Novak (1992) through various numerical benchmarks. The results validate our theoretical findings and demonstrate our methods' advantages for different monotone function classes. Additionally, we include a QMRA application case benchmark, using a simplified version of the actual integration problem, to underscore the benefits of monotone integration methods.

Multiobjective Optimization. The last chapter, Chapter 4, is devoted to studying multiobjective optimization algorithms for computationally intensive stochastic simulators, referred to as Multi Objective Simulation Optimization (MOSO) problems (Hunter et al., 2019). This investigation is driven by the Arti-SaneFood project's goal to optimize the QMRA model's process intervention parameters. The chapter reviews existing algorithms within the Bayesian design and analysis framework for MOSO problems. We examine two primary approaches for creating a sampling criterion or acquisition function for evaluation points: 1) "maximal uncertainty sampling" and 2) "Stepwise Uncertainty Reduction (SUR)." In the maximal uncertainty sampling context, our study reveals that some well-known sampling criteria from reliability theory literature are inefficient for MOSO problems, indicating that not all uncertainty measures are suitable as sampling criteria. However, algorithms based on the weighted Mean Squared Error (w-MSE) measure show promise. We benchmark w-MSE-based algorithms against the PALS (Pareto Active Learning for Stochastic simulators) algorithm Barracosa et al. (2021), but our results confirm that w-MSE variants do not consistently outperform PALS. This led us to explore SUR-based strategies (Vazquez and Bect, 2009, Vazquez and Martinez, 2006, Villemonteix et al., 2007) for MOSO problems. We introduce a new algorithm based on the weighted-Integrated Mean Squared Error (w-IMSE) (Basak et al., 2023a), inspired by the SUR principle.

Numerical benchmarks demonstrate significant improvements over PALS in terms of simulation budget needed to achieve a specified performance level. However, our findings also affirm that PALS remains a viable, cost-effective, and straightforward w-MSE algorithm for MOSO problems. Lastly, Chapter 4 proposes an adaptation of the PALS algorithm Basak et al. (2022a) for the QMRA application, showcasing its application in optimizing intervention parameters for the single-pathogen QMRA model of MPS STEC.

5.2 . Limitations and Future Work

The research conducted in this thesis lays the groundwork for several potential future research directions.

Quantitative Risk Assessment Model. Firstly, while QMRA models offer mathematical representations of microbiological systems, they inevitably remain incomplete due to their inherent complexity and various unaccounted factors (World Health Organization et al., 2021). Validating these models is challenging; however, data-driven methods and comparisons with existing models in literature, as outlined in Section 2.5.2 of Chapter 2, provide a framework for this. The developed multipathogen model builds upon various QMRA literature sources (especially, Bonifait et al., 2021, Perrin et al., 2014, Sanaa

et al., 2004, Strickland et al., 2023) and expert input from ASNES, CNIEL, and ACTALIA. Nonetheless, the model presents opportunities for enhancement.

An improvement area is refining the preharvest intervention step to align with actual farm milk testing protocols used by cheese producers. Typically, a farm with milk contamination is excluded from the production chain until it meets standard hygiene norms, often involving additional costs for enhanced farm inspections. A potential modification could involve updating the farm hygiene parameters (namely, α and σ^2) for each simulated batch. However, this change would create dependencies between batches, increasing simulation costs and limiting parallel computation capabilities.

The proposed DALY metric hinges on critical assumptions, prompting the need for further exploration. This involves a comprehensive literature review to gather relevant epidemiological data and the estimation of age-dependent DALY metrics specific to cheese-borne illnesses in the context of France.

Furthermore, the model could benefit from the enhancements discussed in Section 2.5.1, including 1) incorporating milk testing results for all pathogens in the preharvest intervention; 2) focusing postharvest intervention on highly virulent pathogen strains; and 3) implementing a more dynamic and realistic secondary growth model for *Listeria monocytogenes*.

Integration of Bounded Monotone Functions. As mentioned in Section 3.6, two primary limitations exist when applying monotone bounded integration methods to the QMRA simulator: 1) the presence of internal stochastic variables disrupting the monotonicity of simulator outputs, and 2) the unknown distribution of the integrating variable, hindering domain-specific integration reduction. For the first challenge, several approaches have been discussed to achieve monotonicity in the QMRA simulator implementation. The second limitation, common in industrial problems involving stochastic simulators, can be addressed by examining the cost trade-off between sampling from the integrating variable and the simulator. If sampling the integrating variable is significantly less expensive than simulating the outputs, monotone bounded integration methods could be advantageous.

A future research direction is to explore sequential algorithms based on "greedy" strategies for first-level strata determination, focusing on strata bound differences. These strategies could be numerically efficient for step-like functions, a key interest area, compared to equispaced strata approaches. However, establishing worst-case upper bounds for such methods poses a significant challenge.

Multiobjective Optimization. Our study in Chapter 3 concentrates on developing a sampling criterion for Bayesian optimization algorithms. We primarily focused on the Volume of Symmetric Difference (VSD) metric, which

assesses the algorithm's performance on the Pareto front, indicating accuracy in the objective space. However, other metrics like the Misclassification Rate (MCR), which evaluates performance on the Pareto set (i.e., accuracy in the input space), also exist. The weighted-Integrated Mean Squared Error (w-IMSE) method we proposed bases its weights on the dominated area in the objective space, making it effective for estimating the Pareto front and consistently outperforming PALS on the VSD metric. Yet, the w-IMSE method did not consistently excel over PALS in terms of the MCR metric. Other w-IMSE variants with a partial focus on the misclassification rate were tested against PALS, but none proved consistently superior in both VSD and MCR metrics. Thus, developing algorithms that enhance performance on both the Pareto front and set remains a potential area for future research. Our findings indicate that global uncertainty strategies, such as SUR or w-IMSE methods, often outperform maximal uncertainty sampling methods. However, w-MSE based algorithms like PALS (Barracosa et al., 2021) can also be highly efficient.

In the QMRA application context, some objectives are not direct simulator outputs. Hence, we proposed an extended version of PALS Basak et al. (2022a) suitable for this framework. However, for the w-IMSE method, the sampling criterion is not analytically tractable in this context and requires approximation. Additionally, this criterion is currently defined only for a two-objective framework, based on the dominated area in the objective space. A natural progression would be to expand the w-IMSE algorithm for multiobjective setups and nonlinear outputs.

Our research focused on w-IMSE methods due to their analytically tractable sampling criterion, unlike most SUR-based strategies that necessitate approximating the criterion via Monte Carlo simulations. An interesting future direction would be to devise analytically tractable SUR strategies within the MOSO framework, eliminating the need for criterion approximation. Another significant aspect of this work involves expanding the numerical benchmark to compare the w-IMSE method and PALS against other methods in the literature (Belakaria et al., 2020, Hernández-Lobato et al., 2014) for MOSO problems.

Synthèse

Ce travail de thèse s'inscrit dans le domaine de la modélisation probabiliste et de la quantification de l'incertitude, dans le cadre du projet européen ArtiSaneFood soutenu par le programme PRIMA de l'Union européenne. Il vise à développer des interventions biologiques innovantes et des approches de modélisation des risques pour garantir la sécurité microbienne et la qualité des aliments fermentés artisanaux méditerranéens. Au niveau national en France, le projet aborde un problème industriel lié à l'évaluation du risque microbiologique dans l'industrie laitière, en particulier dans la production de fromage à pâte molle au lait cru, dans le but d'optimiser les stratégies d'intervention. Plus précisément, le problème est formulé comme une optimisation bayésienne d'un modèle d'appréciation quantitative du risque microbiologique (AQRM) afin d'établir des stratégies d'intervention biologique efficaces liées aux tests de lait ou de fromage pour les producteurs de fromage en France. Le travail est structuré en trois directions principales : 1) développement et mise en œuvre d'un modèle AQRM multipathogène pour le fromage à pâte molle au lait cru, 2) étude des méthodes d'intégration numérique pour les fonctions monotones et bornées afin d'estimer les sorties du modèle AQRM, et 3) conception d'un algorithme d'optimisation bayésienne adapté à un simulateur AQRM stochastique et coûteux en calcul.

Le premier chapitre de la thèse porte sur le développement d'un modèle mathématique permettant d'évaluer le risque de maladies d'origine alimentaire et le coût des interventions associées à la production de fromage à pâte molle au lait cru. Ce travail propose le premier modèle AQRM (Appréciation Quantitative du Risque Microbiologique) multipathogène (Basak et al., 2023b) pour le fromage à pâte molle au lait cru, s'appuyant sur des études existantes dans la littérature (voir, par exemple, Bonifait et al., 2021, Perrin et al., 2014, Sanaa et al., 2004, Strickland et al., 2023). Le modèle proposé évalue l'impact des maladies d'origine alimentaire sur la santé publique causé par trois classes de pathogènes potentiellement présents dans le fromage à pâte molle au lait cru : Main Pathogenic Serotypes of Shiga Toxin producing Escherichia coli (MPS-STEC), *Salmonella* et *Listeria monocytogenes*. Ce modèle, du ferme à la fourchette, reproduit chaque étape de la production de fromage pour estimer le risque de maladies d'origine alimentaire en termes de l'espérance de vie corrigée de l'incapacité (EVCI) résultant de la consommation de fromage au lait cru. En d'autres termes, le modèle simule mathématiquement chaque étape, de la collecte du lait à la ferme au traitement dans l'usine de production de fromage et à la consommation par la population, en tenant compte des habitudes de consommation de fromage. De plus, le modèle AQRM intègre des stratégies d'intervention liées aux tests de lait et

de fromage pour estimer les coûts des interventions. Une implémentation du modèle AQRM pour les STEC est fournie en R et dans le cadre du FSKX (Basak et al., 2024).

Le deuxième chapitre de ce manuscrit explore les méthodes d'intégration numérique, à la fois séquentielles et non séquentielles, en explorant leur application potentielle en exploitant les propriétés de monotonie et de bornage des sorties du simulateur AQRM. Il propose une revue exhaustive de la littérature sur les méthodes d'intégration existantes (voir, par exemple, Kiefer, 1957, Novak, 1992), en mettant principalement l'accent sur les méthodes aléatoires non séquentielles et séquentielles, ainsi que leur convergence théorique. La contribution principale comprend un examen détaillé de la méthode aléatoire séquentielle proposée par Novak (1992) et des propositions d'améliorations, telles que la dérivation d'une borne supérieure plus serrée de la variance de l'estimateur, l'amélioration du schéma d'allocation budgétaire par strate et la sélection du nombre initial de strates. Une série d'expériences de benchmark numérique évalue les performances de ces améliorations, démontrant des avantages significatifs pour l'intégration de fonctions en escalier ou abruptes, typiques dans le cas d'application AQRM. Le chapitre se conclut par une discussion des défis associés à l'application de ces méthodes d'intégration dans le cadre AQRM et fournit un exemple d'application sur une version simplifiée du problème.

Le dernier chapitre de ce manuscrit étudie les algorithmes d'optimisation bayésienne adaptés à l'optimisation de simulateurs stochastiques et computationnellement coûteux. Cette étude est motivée par l'application AQRM où le décideur est intéressé à trouver tous les paramètres d'entrée optimaux de modèle AQRM qui minimisent simultanément les sorties du simulateur AQRM, à savoir le risque de maladies d'origine alimentaire mesuré en termes d'EVCI et le coût des interventions. Ce chapitre étudie deux grandes catégories d'algorithmes dans le cadre de la conception séquentielle bayésienne d'expériences informatiques, à savoir 1) Maximal Uncertainty Sampling (MUS) and 2) Stepwise Uncertainty Reduction (SUR). Dans le contexte de MUS, l'étude révèle que certains critères d'échantillonnage bien connus de la littérature sur la théorie de la fiabilité sont inefficaces pour les problèmes de MOSO (Multi-Objective Simulation Optimization), ce qui indique que toutes les mesures d'incertitude ne conviennent pas comme critères d'échantillonnage. Cependant, les algorithmes basés sur la mesure de l'erreur quadratique moyenne pondérée (w-MSE) montrent des promesses. Un benchmark numérique pour les algorithmes basés sur w-MSE contre l'algorithme PALS (Pareto Active Learning for Stochastic simulators) Barracosa et al. (2021) est fourni, mais les résultats confirment que les variantes w-MSE ne surpassent pas systématiquement PALS. Cela a conduit à l'exploration de stratégies basées sur SUR (voir, par exemple, Vazquez and Bect, 2009, Vazquez and Martinez, 2006, Villemon-

teix et al., 2007) pour les problèmes MOSO. Ce chapitre introduit un nouvel algorithme basé sur l'erreur quadratique moyenne intégrée pondérée (w-IMSE) (Basak et al., 2023a), inspiré par le principe de SUR. Les benchmarks numériques démontrent des améliorations significatives par rapport à PALS en termes de budget de simulation nécessaire pour atteindre un niveau de performance spécifié. Cependant, les résultats confirment également que PALS reste un algorithme w-MSE viable, rentable et simple pour les problèmes MOSO. Le chapitre se termine par une adaptation de l'algorithme PALS (Basak et al., 2022a) pour l'application AQRM, illustrant son application dans l'optimisation des paramètres d'intervention pour le modèle AQRM monopathogène de MPS STEC.

Appendices

A - Estimation of farm hygiene parameters

In Perrin et al. (2014) the authors proposed a hierarchical Poisson mixed model to express the relationship between the hygiene parameters α and σ^2 for each farm, that determines the *E. coli* concentration, denoted by x_d, in BTM on day d,

$$\begin{aligned} x_d \mid \Lambda_d &\sim \text{Poisson}(\Lambda_d) \\ \log(\Lambda_d) &= \alpha + \varepsilon_d, \text{ where } \varepsilon_d \sim \mathcal{N}(0, \sigma^2), \end{aligned} \tag{A.1}$$

where Λ_d denotes the expected *E. coli* concentration on day d in a particular farm. *E. coli* test data was collected under the ArtiSaneFood project work packages, from three different French cheese producers, namely site A, B and C, over the years $2019 - 2022$, covering a total of 104 farms. This data corresponds to the tests performed at the farm level to quantify the concentration of *E. coli* in the BTM of the farms. The number of farms under each site and the corresponding number of tests are shown in Table A.1.

Table A.1: Specifications of *E. coli* test data.

Site name	A	B	C	Total
Number of farms	44	39	21	104
Number of tests	1057	1508	6310	8875

These 8875 tests consist of two types, namely "routine" and "renforcée", and for the estimation of hygiene parameters only "routine" tests as considered. The recorded *E. coli* concentration on day d, denoted by $x_{0,d}$, was lower truncated for observations less than 10 CFU/mL and upper truncated for observations above 150 CFU/mL for site B, or 300 CFU/mL for site A and C, as demonstrated in Table A.2.

Table A.2: Specifications of "routine" test data.

	< 10	> 150	> 300	Exact count	Total
A	940	0	20	87	1047
B	1353	17	0	66	1436
C	6068	0	54	188	6310
Total	8361	17	74	341	8793

We propose a Bayesian MCMC sampling approach, to estimate the values of α and σ^2 for each farm. First we define an alternate parameterization of

the model by applying exponential on (A.1) : $\Lambda_d = \bar{\Lambda} \cdot \eta_d$, where $\bar{\Lambda} = \exp(\alpha)$ and $\eta_d = \exp(\varepsilon_d)$. Further define $\underline{x} = \{x_1, x_2, ..., x_D\}$ and $\underline{\eta} = \{\eta_1, \eta_2, ..., \eta_D\}$, where D is the total number of tests for that particular farm. Corresponding to each x_d, the variable c_d denotes the truncation class: -1 for lower truncation, 1 for upper truncation and 0 for no truncation. We choose conditionally conjugate, non-informative priors $\bar{\Lambda} \sim \Gamma(a_{\bar{\Lambda}}, b_{\bar{\Lambda}})$ and $\sigma^2 \sim \text{InvGamma}(a_{\sigma^2}, b_{\sigma^2})$, such that the posterior distributions are derived in closed forms of known distributions.

$$\begin{aligned} p(\bar{\Lambda}|\underline{x}, \underline{\eta}) &\propto p(\underline{x}|\bar{\Lambda}, \underline{\eta}) p(\bar{\Lambda}|\underline{\eta}) \\ &= \prod_{d=1}^{D} e^{-(\bar{\Lambda}\eta_d)} \frac{(\bar{\Lambda}\eta_d)^{x_d}}{x_d!} \times \frac{b_{\bar{\Lambda}}^{a_{\bar{\Lambda}}}}{\Gamma a_{\bar{\Lambda}}} \bar{\Lambda}^{(a_{\bar{\Lambda}}-1)} e^{-b_{\bar{\Lambda}}\bar{\Lambda}} \\ &\propto \bar{\Lambda}^{(\sum x_d + a_{\bar{\Lambda}} - 1)} e^{-\bar{\Lambda}(b_{\bar{\Lambda}} + \sum \eta_d)} \end{aligned} \tag{A.2}$$

$$\begin{aligned} p(\sigma^2|\underline{\eta}) &\propto p(\underline{\eta}|\sigma^2) p(\sigma^2) \\ &= \prod_{d=1}^{D} \frac{1}{\sigma\sqrt{2\pi}\eta_d} e^{-\frac{1}{2\sigma^2} log\eta_d^2} \times \frac{b_{\bar{\sigma^2}}^{a_{\bar{\sigma^2}}}}{\Gamma a_{\bar{\sigma^2}}} \sigma^{-2(a_{\bar{\sigma^2}}+1)} e^{-b_{\bar{\sigma^2}}/\sigma^2} \end{aligned} \tag{A.3}$$

The posterior distributions for $\bar{\Lambda}$ and σ^2 are obtained as,

$$\begin{aligned} (\bar{\Lambda}|\underline{x}, \underline{\eta}) &\sim \Gamma(a_{\bar{\Lambda}} + \sum x_d, b_{\bar{\Lambda}} + \sum \eta_d) \\ (\sigma^2|\underline{\eta}) &\sim \text{InvGamma}(a_{\sigma^2} + \frac{D}{2}, b_{\sigma^2} + \frac{1}{2}\sum_{d=1}^{D} (\log(\eta_d))^2) \end{aligned} \tag{A.4}$$

Now to sample form the target posteriors we use a Gibbs sampler with a Metropolis-Hastings step for the conditional distribution of η_d. The iterative algorithm starts with initialization of the variables $\{c_d^0, \underline{x}^0, \underline{\eta}^0, \bar{\Lambda}^0, \sigma^{2,0}\}$. The steps of the t-th iteration, $t = 1, 2, \ldots, T$, are described below.

Step 1. Assuming η_ds are independently distributed, sample η_d^t separately from the conditional distribution $(\eta_d|x_d^{t-1}, \bar{\Lambda}^{t-1}, \sigma^{2,t-1})$, for each $d = 1, ..., D$, using Random Walk Metropolis-Hastings (RW-MH) algorithm. For the k-th step, $k = 1, ..., K$, of the MH algorithm, we use a LogNormal proposal density with log mean at η_d^{k-1} and log standard deviation $\log(\sigma^{t-1})$. The un-normalized target density used in the MH algorithm, is of the following form,

$$p(\eta_d|x_d, \bar{\Lambda}, \sigma^2) \propto e^{-(\bar{\Lambda}\cdot\eta_d)} \frac{(\bar{\Lambda} \cdot \eta_d)^{x_d}}{x_d!} \times \frac{1}{\sigma\eta_d\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2} log(\eta_d)^2} \tag{A.5}$$

Step 2. Sample $\bar{\Lambda}^t$ from the conditional distribution $(\bar{\Lambda}|\underline{x}^{t-1}, \underline{\eta}^t)$, given by,

$$(\bar{\Lambda}|\underline{x}^{t-1}, \underline{\eta}^t) \sim \Gamma(a_{\bar{\Lambda}} + \sum x_d^{t-1}, b_{\bar{\Lambda}} + \sum \eta_d^t) \tag{A.6}$$

Step 3. Sample $\sigma^{2,t}$ from the conditional distribution $(\sigma^2|\underline{\eta}^t)$, given by,

$$(\sigma^2|\underline{\eta}^t) \sim \text{InvGamma}(a_{\sigma^2} + \frac{D}{2}, b_{\sigma^2} + \frac{1}{2}\sum_{d=1}^{D}(\log(\eta_d^t)^2) \tag{A.7}$$

Step 4. Sample x_d^t (for $d = 1, 2, ..., D$) from the conditional distribution $(x_d|x_{0,d}, \eta_d^t, c_d^{t-1}, \bar{\Lambda}^t)$, given by,

$$(x_d|x_{0,d}, \eta_d^t, c_d^{t-1}, \bar{\Lambda}^t) \sim \text{Poisson}(\eta_d^t \cdot \bar{\Lambda}^t) \tag{A.8}$$

given $x_d \in c_d^t$

These steps of the Gibbs sampler are repeated iteratively to simulate the posterior distributions of the hygiene parameters, with the corresponding posterior means as their estimates. We used $T = 5000$ iterations of the Gibbs sampler with $K = 5$ for the MH step. The parameters of the non-informative prior distributions are taken as $a_{\bar{\Lambda}} = 1$, $b_{\bar{\Lambda}} = 1$, $a_{\sigma^2} = 1$ and $b_{\sigma^2} = 1$. The initial parameter values were chosen adaptively by another pilot run of the Gibbs sampler, with $\bar{\Lambda}^0 = \exp(0.85)$, $\sigma^0 = 0.8$ and a burnin rate of 50%.

B - User inputs of the FSKX model

The FSKX implementation of the QMRA model for STEC allows the user to run the model with the default input values listed in the following table.

Table B.1: Default simulation settings.

Symbol	Values
fm_N_farms	31
fm_q_milk	25
fm_sorting_freq	10
fm_sorting_lim	50
fm_mu_u	-0.927
fm_tau_u	1.47411
fm_a_weibull	0.264
fm_b_weibull	16.288
fm_mu_ecoli	6
fm_tau_ecoli	0.3
cm_mu_max_T_min	5.5
cm_mu_max_T_opt	40.6
cm_mu_max_T_max	48.1
cm_mu_max_pH_min	3.9
cm_mu_max_pH_opt	6.25
cm_mu_max_pH_max	14
cm_mu_max_aw_min	0.9533
cm_mu_max_aw_opt	0.999
cm_mu_max_mu_opt	2.03
cm_w_activity	0.99
cm_rho_O157H7	0.14
cm_rho_otherMPS	0.033
cm_y_max_milk	$1e+09$
cm_y_max_cheese	$1e+05$
cm_storage_duration	12
cm_storage_duration_min	1
cm_storage_duration_max	40
cm_storage_duration_mode	12
cm_storage_temperature	5
cm_storage_temperature_min	1
cm_storage_temperature_max	6
cm_p_O157H7	0.76
cm_p_MPS_STEC or fm_p_MPS_STEC	0.025
cm_mu_eps_O157H7	0
cm_tau_eps_O157H7	0.000279659

Continued on the next page

Symbol	Values
cm_mu_eps_otherMPS	0
cm_tau_eps_otherMPS	$6.5399e-05$
cm_molding_duration	3
cm_draining_duration	17
cm_salting_duration	4.5
cm_consumption_time_min	22
cm_consumption_time_max	60
cm_consumption_time_mode	30
cm_v_cheese	2200
cm_w_loss	0.9
cm_wt_cheese	250
cm_wt_serving	25
cm_m_sample	25
cm_n_sample	5
cm_k	0.38
cm_ro	$1e-2.33$
cm_age_max	14
cm_p_test	0.5
cm_d_test	14
cm_n_dose	0
cm_n_batch	1
flag_consum	TRUE
flag_MPS	FALSE

C - Performance of w-MSE and w-IMSE algorithms

This section presents the additional results from the numerical benchmark experiments with w-MSE and w-IMSE methods, on the test problems from Barracosa et al. (2021). Figure C.1, Figure C.2 and Figure C.3 shows the performance of the w-MSE algorithm with the proposed weights (w-MSE-vsd) compared to the propsoed w-IMSE algorithm, PALS and the Random search algorithm. Although the w-MSE-vsd method showed a slight improvement with respect to the VSD metric, it performed very poorly with respect to the MCR metric.

Figure C.4 monitors the performances of several proposed w-MSE methods in terms of the VSD metric, which do not exhibit any consistent improvement over the PALS algorithm.

Figure C.5 shows the performance of the proposed w-IMSE algorithm on the estimation of the Pareto set, monitored by the misclassification rate (MCR) over all the test problems from Barracosa et al. (2021). Clearly the proposed w-IMSE algorithm does not show improvement in terms of the MCR metric compared to the PALS algorithm.

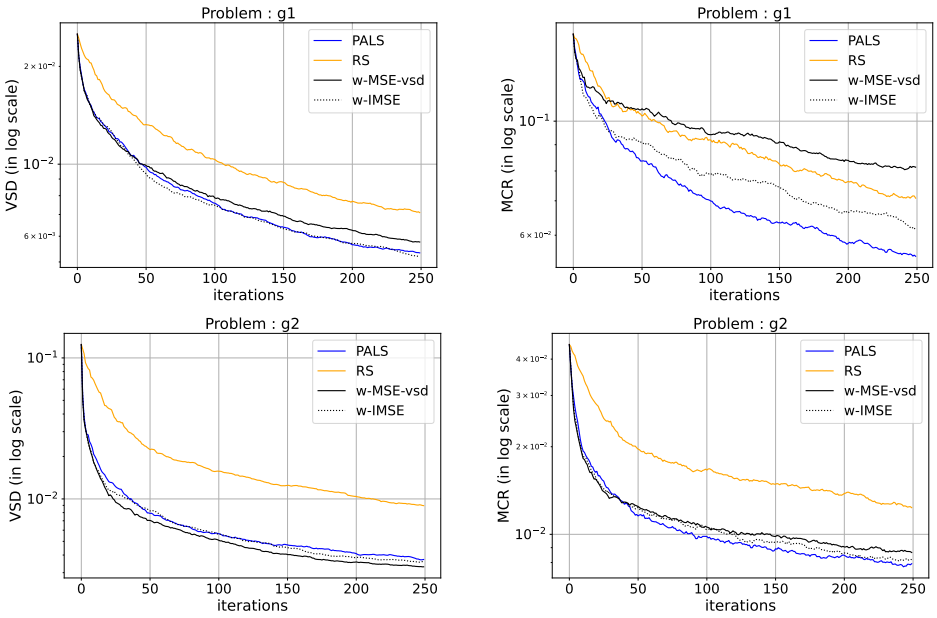


Figure C.1: Average VSD (on left) and MCR (on right) metric comparing the w-MSE-vsd, w-IMSE, PALS and RS methods on test problems $g1$ and $g2$ from Barracosa et al. (2021).

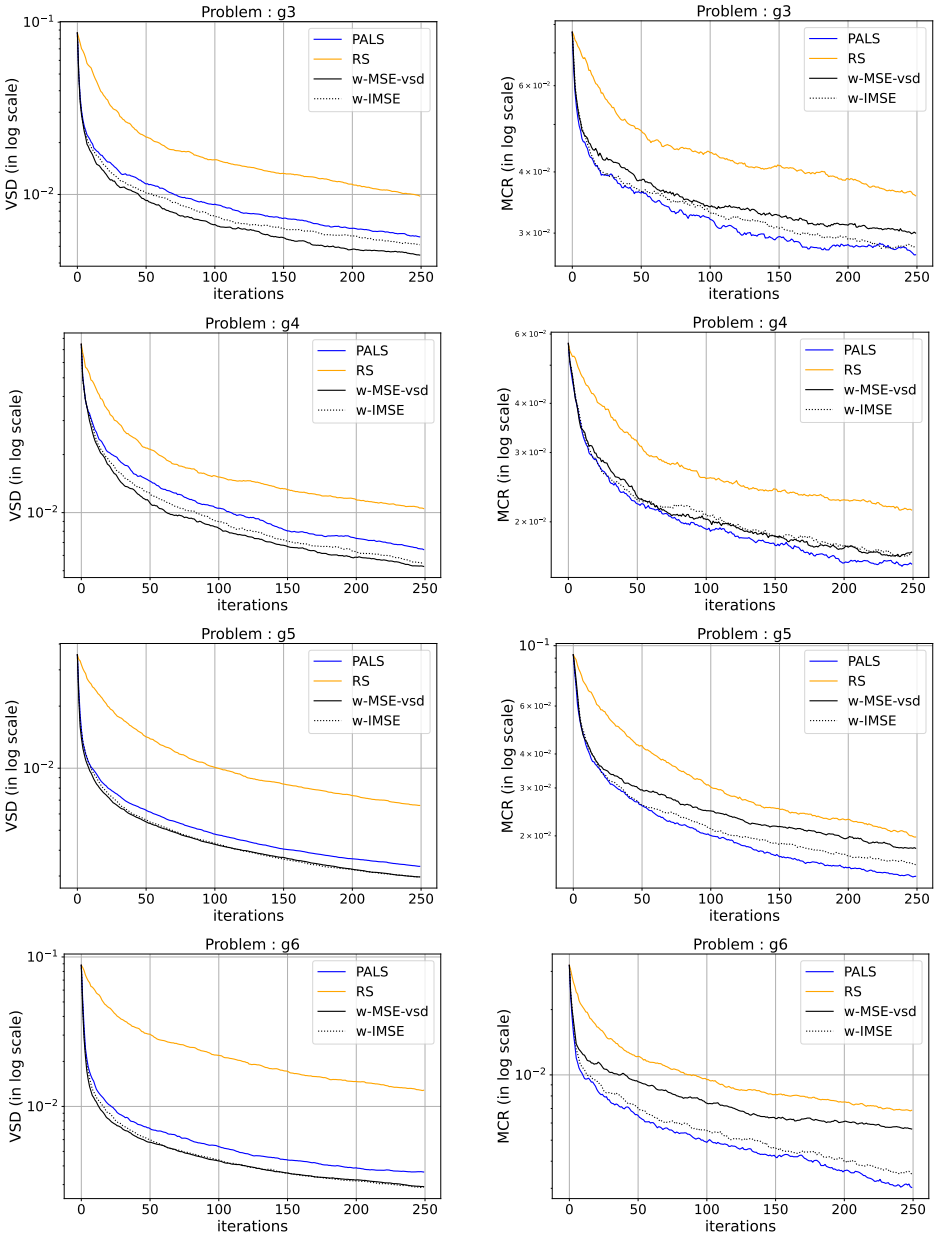


Figure C.2: Average VSD (on left) and MCR (on right) metric comparing the w-MSE-vsd, w-IMSE, PALS and RS methods on test problems $g3 - g6$ from Barracosa et al. (2021).

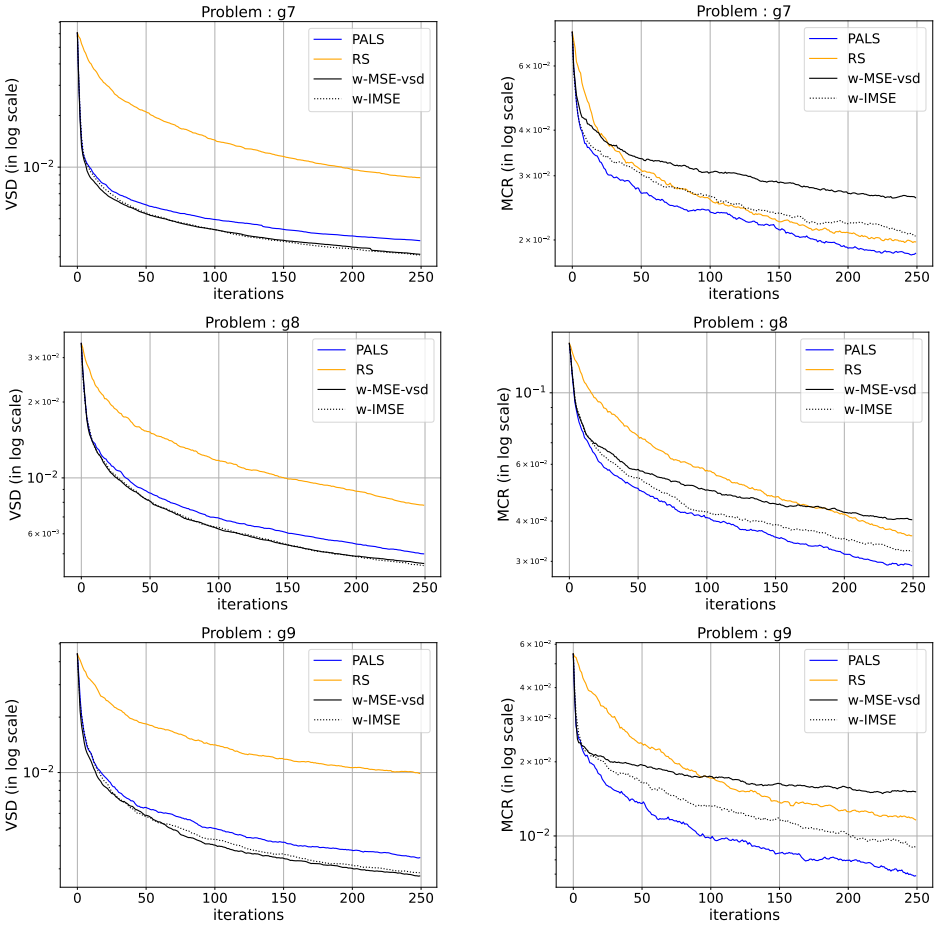


Figure C.3: Average VSD (on left) and MCR (on right) metric comparing the w-MSE-vsd, w-IMSE, PALS and RS methods on test problems $g7 - g9$ from Barracosa et al. (2021).

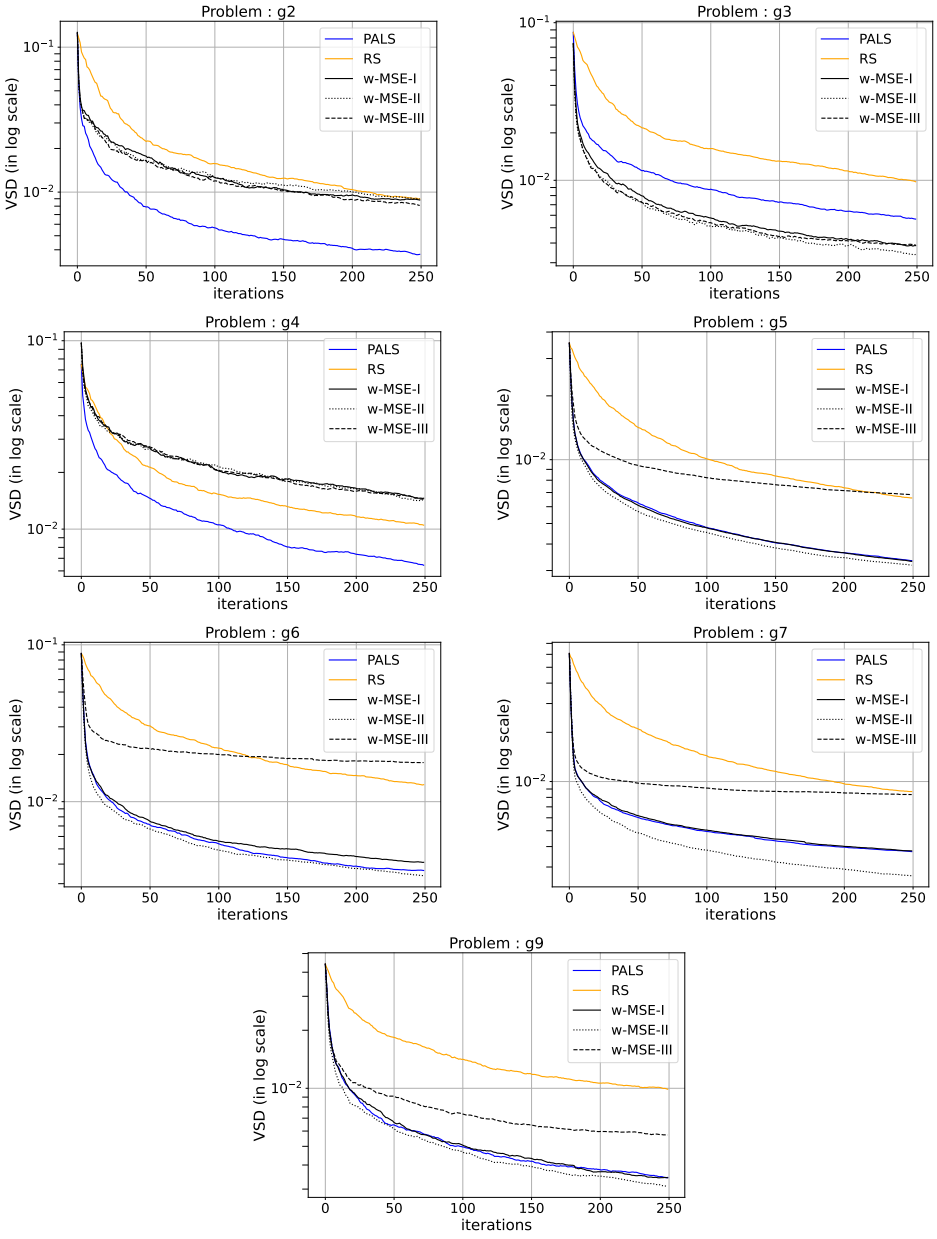


Figure C.4: Average VSD metric comparing the proposed w-MSE methods on test problems $g2 - g7$ and $g9$ from Barracosa et al. (2021).

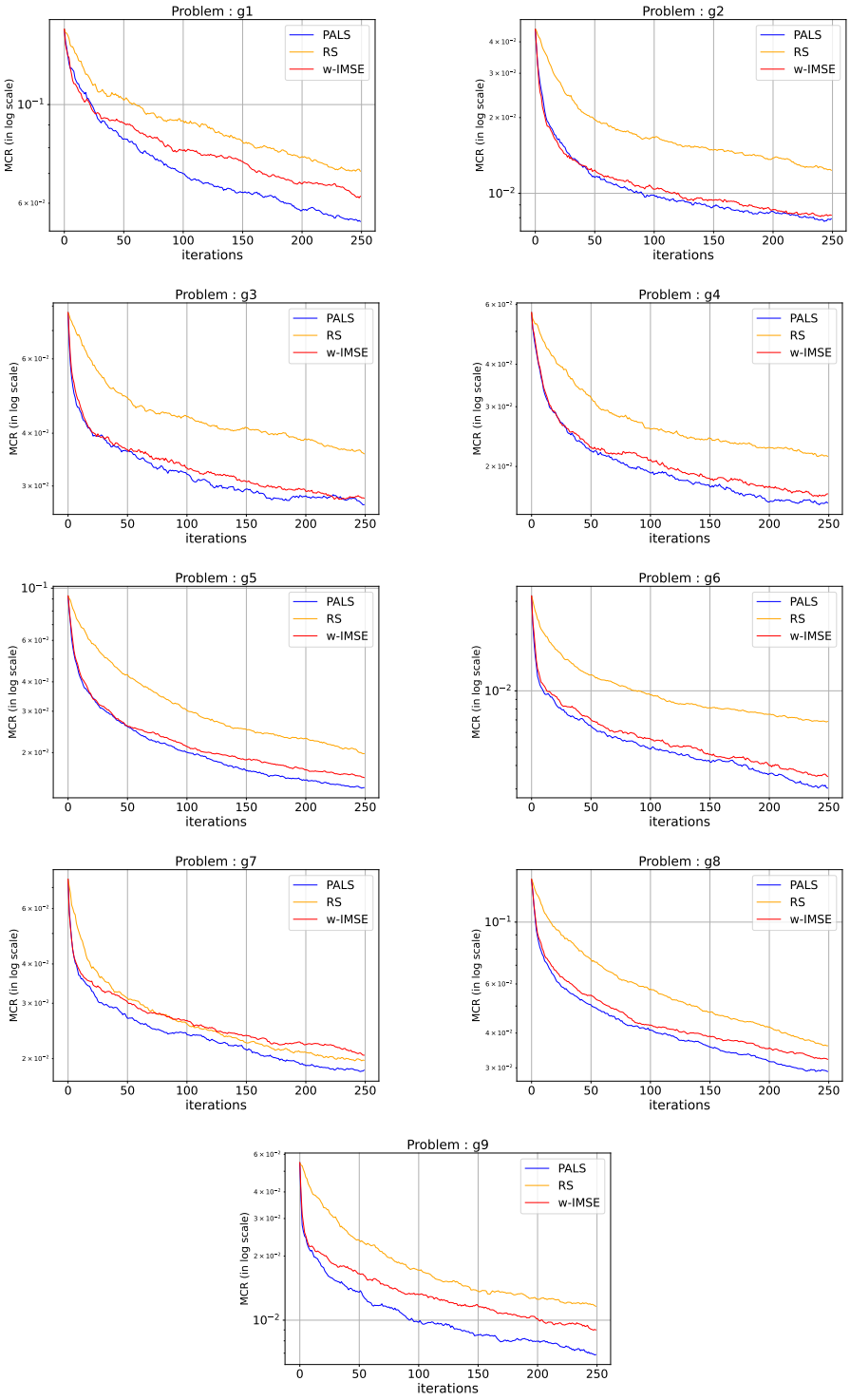


Figure C.5: Average MCR metric of the proposed w-IMSE based algorithm.

Bibliography

R. A. Abdelmalek-Lomenech, J. Bect, V. Chabridon, and E. Vazquez. Bayesian sequential design of computer experiments to estimate reliable sets. *arXiv preprint arXiv:2211.01008*, 2022.

ANSES. Méthodologie de hiérarchisation des dangers biologiques et chimiques dans les aliments. 2020.

R. Astudillo and P. Frazier. Multi-attribute bayesian optimization under utility uncertainty. In *Proceedings of the NIPS Workshop on Bayesian Optimization*, volume 172, 2017.

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.

F. Auvray, A. Cointe, M. Desvaux, G. Jones, P. Mariani Kurkdjian, E. Oswald, P. Kooh, L. Guillier, and F. Audiat-Perrin. Avis de l'Anses relatif à la définition des souches pathogènes d'Escherichia coli productrices de shigatoxines. Technical Report Saisine n°2020-SA-0095, Anses, May 2023. URL https://anses.hal.science/anses-04171498. Citation suggérée : Anses (2023). Avis relatif à la définition des souches pathogènes d'Escherichia coli productrices de shigatoxines (saisine n°2020-SA-0095). Maisons-Alfort : Anses, 63 p.

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, S. Petit, J. Bect, and E. Vazquez. Numerical issues in maximum likelihood parameter estimation for gaussian process interpolation. In *Machine*

Learning, Optimization, and Data Science, pages 116–131. Springer International Publishing, 2022c. ISBN 978-3-030-95470-3.

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: $10.3897/\mathrm{fmj}.5.109502$. URL https://doi.org/10.3897/fmj.5.109502.

J. Bect, F. Bachoc, D. Ginsbourger, et al. A supermartingale approach to gaussian process based sequential design of experiments. *Bernoulli*, 25(4A): 2883–2919, 2019.

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.

M. Binois, J. Huang, R. Gramacy, and M. Ludkovski. Replication or exploration? sequential design for stochastic simulation experiments. *Technometrics*, 61, 10 2017. doi: $10.1080/00401706.2018.1469433$.

M. Binois, R. B. Gramacy, and M. Ludkovski. Practical heteroscedastic gaussian process modeling for large simulation experiments. *Journal of Computational and Graphical Statistics*, 27(4):808–821, 2018.

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.

Hans W. Borchers. *pracma: Practical Numerical Math Functions*, 2022. URL https://CRAN.R-project.org/package=pracma. R package version 2.4.2.

J. Branke and W. Zhang. A new myopic sequential sampling algorithm for multi-objective problems. In *2015 Winter Simulation Conference (WSC)*, pages 3589–3598. IEEE, 2015.

M. Bruyand, P. Mariani-Kurkdjian, S. Le Hello, L. A. King, D. Van Cauteren, S. Lefevre, M. Gouali, N. Jourdan-da Silva, A. Mailles, M. P. Donguy, et al. Paediatric haemolytic uraemic syndrome related to shiga toxin-producing escherichia coli, an overview of 10 years of surveillance in france, 2007 to 2016. *Eurosurveillance*, 24(8):1800068, 2019.

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.

V. Cadavez, R. Pouillot, L. Guillier, U. Gonzales-Barron, and M. Sanaa. *JEMRA QMRA models for Listeria monocytogenes*, unpublished. URL https://github.com/vcadavez/JEMRA/. R package version 0.2.0.

A. C. Camargo, J. J. Woodward, D. R. Call, and L. A. Nero. Listeria monocytogenes in food-processing facilities, food contamination, and human listeriosis: the brazilian scenario. *Foodborne Pathogens and Disease*, 14(11):623–636, 2017.

F. B. Campagnollo, U. Gonzales-Barron, V. A. Pilao Cadavez, A. S. Sant'Ana, and D. W. Schaffner. Quantitative risk assessment of listeria monocytogenes in traditional minas cheeses: The cases of artisanal semi-hard and fresh soft cheeses. *Food Control*, 92:370–379, 2018. ISSN 0956-7135.

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.

J. P. Chiles and P. Delfiner. *Geostatistics: modeling spatial uncertainty*, volume 713. John Wiley & Sons, 2012.

W. G. Cochran. *Sampling Techniques, 3rd Edition.* John Wiley, 1977. ISBN 0-471-16240-X.

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

D. A. Cohn, Z. Ghahramani, and M. I. Jordan. Active learning with statistical models. *Journal of artificial intelligence research*, 4:129–145, 1996.

L. Collineau, B. Chapman, X. Bao, B. Sivapathasundaram, C. A. Carson, A. Fazil, R. J. Reid-Smith, and B. A. Smith. A farm-to-fork quantitative risk assessment model for salmonella heidelberg resistant to third-generation

cephalosporins in broiler chickens in canada. *International journal of food microbiology*, 330:108559, 2020.

N. Commeau. *Modélisation de la contamination par Listeria monocytogenes pour l'amélioration de la surveillance dans les industries agro-alimentaires*. Theses, AgroParisTech, Jun 2012. URL https://pastel.hal.science/pastel-00770790.

N. Costanzo, C. Ceniti, A. Santoro, M. T. Clausi, F. Casalinuovo, et al. Foodborne pathogen assessment in raw milk cheeses. *International Journal of Food Science*, 2020, 2020.

N. Cressie. Aggregation in geostatistical problems. In *Geostatistics Tróia'92: Volume 1*, pages 25–36. Springer, 1993.

C. Currin, T. Mitchell, M. Morris, and D. Ylvisaker. Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*, 86(416):953–963, 1991.

Van C. D., Jourdan da Silva N., F. X. Weill, L. King, A. Brisabois, G. Delmas, V. Vaillant, and H. De Valk. Outbreak of salmonella enterica serotype muenster infections associated with goat's cheese, france, march 2008. *Eurosurveillance*, 14(31):19290, 2009.

E. Dalzini, V. Bernini, B. Bertasi, P. Daminelli, M. N. Losio, and G. Varisco. Survey of prevalence and seasonal variability of listeria monocytogenes in raw cow milk from northern italy. *Food Control*, 60:466–470, 2016.

F. Dubois-Brissonnet, F. Auvray, O. Cerf, C. Delbès, P. Fravalo, M. Gautier, N. Jourdan Da-Silva, N. Korsak Koulagenko, F. Mathieu, E. Oswald, H. E. Spinnler, R. Talon, H. Bergis, C. Bayourthe, S. Duret, N. Ramarao, L. Guillier, F. Audiat-Perrin, E. Chaix, P. Kooh, and M. Sanaa. Avis de l'Anses relatif aux modalités de maîtrise du risque lié à la présence de dangers microbiologiques dans les fromages et autres produits laitiers fabriqués à partir de lait cru. Technical Report Saisine n° 2019-SA-0033, Anses, Jan 2022. URL https://anses.hal.science/anses-03889040. Citation suggérée: Anses (2022). Avis relatif aux modalités de maîtrise du risque lié à la présence de dangers microbiologiques dans les fromages et autres produits laitiers fabriqués à partir de lait cru (saisine 2019-SA-0033). Maisons-Alfort: Anses, 126 p.

S. Duret, L. Guillier, H.-M. Hoang, D. Flick, and O. Laguerre. Identification of the significant factors in food safety using global sensitivity analysis and the accept-and-reject algorithm: application to the cold chain of ham. *International journal of food microbiology*, 180:39–48, 2014.

B. Echard, N. Gayton, and M. Lemaire. Ak-mcs: An active learning reliability method combining kriging and monte carlo simulation. *Structural Safety*, 2011.

EFSA. The european union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2010. *EFSA Journal*, 10(3):2597, 2012.

EFSA. The european union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2016. *EFSA journal*, 15(12), 2017.

EFSA BIOHAZ Panel, K. Koutsoumanis, A. Allende, A. Alvarez-Ordonez, S. Bover-Cid, M. Chemaly, R. Davies, A. De Cesare, L. Herman, F. Hilbert, R. Lindqvist, M. Nauta, L. Peixe, G. Ru, M. Simmons, P. Skandamis, E. Suffredini, C. Jenkins, S. Monteiro P., S. Morabito, T. Niskanen, F. Scheutz, M. T. da Silva Felicio, W. Messens, and D. Bolton. Pathogenicity assessment of shiga toxin-producing escherichia coli (stec) and the public health risk posed by contamination of food with stec. *EFSA Journal*, 18(1):e05967, 2020. doi: https://doi.org/10.2903/j.efsa.2020.5967. URL https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2020.5967.

X. Emery. The kriging update equations and their application to the selection of neighboring data. *Computational Geosciences*, 13:269–280, 09 2009. doi: 10.1007/s10596-008-9116-8.

European Centre for Disease Prevention and Control. ECDC BCoDE toolkit, 2019. URL https://ecdc.europa.eu/en/toolkit-application-calculate-dalys.

European Food Safety Authority and European Centre for Disease Prevention & Control. The european union one health 2021 zoonoses report. *EFSA Journal*, 20(12):e07666, 2022.

C. Farrokh, K. Jordan, F. Auvray, K. Glass, H. Oppegaard, S. Raynaud, D. Thevenot, R. Condron, K. De Reu, A. Govaris, et al. Review of shiga-toxin-producing escherichia coli (stec) and their significance in dairy production. *International journal of food microbiology*, 162(2):190–212, 2013.

Joint Institute for Food Safety FDA/CFSAN, Applied Nutrition (JIFSAN), and Risk Sciences International (RSI). Fda-irisk® version 4.2, 2021. URL https://irisk.foodrisk.org/. FDA CFSAN. College Park, Maryland.

Food and Drug Administration et al. Joint fda/health canada quantitative assessment of the risk of listeriosis from soft ripened cheese consumption in

the united states and canada. *Center for Food Safety and Applied Nutrition, US Department of Health and Human Services. Retrieved July*, 15:2014, 2012.

C. Frank, D. Werber, J. P. Cramer, M. Askar, M. Faber, M. an der Heiden, H. Bernard, A. Fruth, R. Prager, A. Spode, et al. Epidemic profile of shiga-toxin–producing escherichia coli o104: H4 outbreak in germany. *New England Journal of Medicine*, 365(19):1771–1780, 2011.

P. Frazier, W. Powell, and S. Dayanik. The knowledge-gradient policy for correlated normal beliefs. *INFORMS Journal on Computing*, 21:599–613, 11 2009. doi: 10.1287/ijoc.1080.0314.

L. Fritsch, L. Guillier, and J. C. Augustin. Next generation quantitative microbiological risk assessment: refinement of the cold smoked salmon-related listeriosis risk model by integrating genomic data. *Microbial Risk Analysis*, 10:20–27, 2018.

D. Geman and B. Jedynak. An active testing model for tracking roads in satellite images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(1):1–14, 1996.

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.

S. R. Gonzalez and I. V. Nieuwenhuyse. A survey on kriging-based infill algorithms for multiobjective simulation optimization. *Computers & Operations Research*, 116:104869, 2020.

S. R. Gonzalez, J. Branke, and I. V. Nieuwenhuyse. Multiobjective ranking and selection using stochastic kriging, 2022.

N. Gopal, C. Hill, P. R. Ross, T. P. Beresford, M. A. Fenelon, and P. D. Cotter. The prevalence and control of bacillus and related spore-forming bacteria in the dairy industry. *Frontiers in microbiology*, 6:1418, 2015.

J. M. Hammersley and D. C. Handscomb. *Monte Carlo methods*. Methuen London, 1964.

E. Hemingway. *A Moveable Feast*. Charles Scribner's Sons, New York, 1964. ISBN PS3515.E37 Z525 1964. URL https://search.library.wisc.edu/catalog/999466397402121.

D. Hernandez-Lobato, J. Hernandez-Lobato, A. Shah, and R. Adams. Predictive entropy search for multi-objective bayesian optimization. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1492–1501, New York, New York, USA, 20–22 Jun 2016. PMLR.

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.

D. Horn, M. Dagge, X. Sun, and B. Bischl. First investigations on noisy model-based multi-objective optimization. In *Evolutionary Multi-Criterion Optimization: 9th International Conference, EMO 2017, Münster, Germany, March 19-22, 2017, Proceedings 9*, pages 298–313. Springer, 2017.

S. R. Hunter, E. A. Applegate, V. Arora, B. Chong, K. Cooper, O. Rincón-Guevara, and C. Vivas-Valencia. An introduction to multi-objective simulation optimization. *ACM Transactions on Modeling and Computer Simulation*, 29(1):7:1–7:36, January 2019. doi: 10.1145/3299872.

H. Jalali, I. Nieuwenhuyse, and V. Picheny. Comparison of kriging-based algorithms for simulation optimization with heterogeneous noise. *European Journal of Operational Research*, 01 2017. doi: 10.1016/j.ejor.2017.01.035.

D. Jones, M. Schonlau, and W. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13:455–492, 12 1998. doi: 10.1023/A:1008306431147.

G. Jones, S. Lefèvre, M. P. Donguy, A. Nisavanh, G. Terpant, E. Fougère, E. Vaissière, A. Guinard, A. Mailles, H. de Valk, et al. Outbreak of shiga toxin-producing escherichia coli (stec) o26 paediatric haemolytic uraemic syndrome (hus) cases associated with the consumption of soft raw cow's milk cheeses, france, march to may 2019. *Eurosurveillance*, 24(22):1900305, 2019.

H. Kahn. *Applications of Monte Carlo*. RAND Corporation, Santa Monica, CA, 1956.

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. A. King, E. Espié, S. Haeghebaert, F. Grimont, P. Mariani-Kurkdjian, I. Filliol-Toutain, E. Bingen, F. X. Weill, C. Loirat, H. De Valk, et al. Surveillance du syndrome hémolytique et urémique chez les enfants de 15 ans et moins en france, 1996-2007. *Bull Epidémiol Hebd*, 14:125–8, 2009.

K. P. Koutsoumanis and Z. Aspridou. Moving towards a risk-based food safety management. *Current Opinion in Food Science*, 12:36–41, 2016.

M. Kretzschmar, M. J. J Mangen, P. Pinheiro, B. Jahn, E. M. Fevre, S. Longhi, T. Lai, A. H. Havelaar, C. Stein, A. Cassini, et al. New methodology for estimating the burden of infectious diseases in europe. *PLoS medicine*, 9(4): e1001205, 2012.

M. Lamboni, M. Sanaa, and F. Tenenhaus-Aziza. Sensitivity analysis for critical control points determination and uncertainty analysis to link fso and process criteria: Application to listeria monocytogenes in soft cheese made from pasteurized milk. *Risk analysis*, 34(4):751–764, 2014.

A. Leclercq, C. Charlier, and M. Lecuit. Global burden of listeriosis: the tip of the iceberg. *The Lancet Infectious Diseases*, 14(11):1027–1028, 2014.

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.

R. Lindqvist, S. Sylvén, and I. Vågsholm. Quantitative microbial risk assessment exemplified by staphylococcus aureus in unripened cheese made from raw milk. *International Journal of Food Microbiology*, 78(1-2):155–170, 2002.

R. Lindqvist, C. Flink, and M. Lindblad. Classification and ranking of shigatoxin-producing escherichia coli (stec) genotypes detected in food based on potential public health impact using clinical data. *Microbial Risk Analysis*, 23: 100246, 2023.

D. J. C. MacKay. Bayesian interpolation. *Neural computation*, 4(3):415–447, 1992.

M. M. Maher, K. N. Jordan, M. E. Upton, and A. Coffey. Growth and survival of e. coli o157: H7 during the manufacture and ripening of a smear-ripened cheese produced from raw milk. *Journal of Applied Microbiology*, 90(2):201–207, 2001.

M. J. J. Mangen, D. Plass, A. H. Havelaar, C. L. Gibbons, A. Cassini, N. Mühlberger, A. van Lier, J. A. Haagsma, R. J. Brooke, T. Lai, et al. The

pathogen-and incidence-based daly approach: an appropriated methodology for estimating the burden of infectious diseases. *PloS one*, 8(11):e79740, 2013.

N. B. McCullough and C. W. Elsele. Experimental human salmonellosis: I. pathogenicity of strains of salmonella meleagridis and salmonella anatum obtained from spray-dried whole egg. *The Journal of infectious diseases*, pages 278–289, 1951.

L. McIntyre, L. Wilcott, M. Naus, et al. Listeriosis outbreaks in british columbia, canada, caused by soft ripened cheese contaminated from environmental sources. *BioMed research international*, 2015, 2015.

M. D. McKay, R. J. Beckman, and Conover W. J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

S. D. Miszczycha, N. Bel, P. Gay-Perret, V. Michel, M. C. Montel, and D. Sergentet-Thevenot. Behavior of different shiga toxin-producing escherichia coli serotypes (o26: H11, o103: H2, o145: H28, o157: H7) during the manufacture, ripening, and storage of a white mold cheese. *Journal of Dairy Science*, 99(7):5224–5229, 2016.

J. Močkus. On bayesian methods for seeking the extremum. In G. I. Marchuk, editor, *Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7, 1974*, pages 400–404, Berlin, Heidelberg, 1975. Springer Berlin Heidelberg. ISBN 978-3-540-37497-8.

C. J. L. Murray and A. D. Lopez. Global mortality, disability, and the contribution of risk factors: Global burden of disease study. *The lancet*, 349(9063):1436–1442, 1997.

J. Neyman. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97(4):558–625, 1934.

E. Novak. Quadrature formulas for monotone functions. *Proceedings of the American Mathematical Society*, 115(1):59–68, 1992.

A. O'Hagan. Some Bayesian numerical analysis. In *Bayesian statistics 4: proceedings of the Fourth Valencia International Meeting, April 15-20, 1991*. Oxford University Press, 1992.

R. Pasupathy, S. R. Hunter, N. A. Pujowidianto, L. H. Lee, and C. H. Chen. Stochastically constrained ranking and selection via score. *ACM Trans.*

Model. Comput. Simul., 25(1), aug 2014. ISSN 1049-3301. doi: 10.1145/2630066. URL https://doi.org/10.1145/2630066.

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 ProducingEscherichia coliStrains in raw milk soft cheeses. *Risk Analysis*, 35(1):109–128, 2014.

C. Plaza-Rodriguez, L. Ungaretti Haberbeck, V. Desvignes, P. Dalgaard, M. Sanaa, M. Nauta, M. Filter, and L. Guillier. Towards transparent and consistent exchange of knowledge for improved microbiological food safety. *Current Opinion in Food Science*, 19:129–137, 2018. ISSN 2214-7993. doi: https://doi.org/10.1016/j.cofs.2017.12.002. URL https://www.sciencedirect.com/science/article/pii/S2214799317301029. Food Chemistry and Biochemistry * Food Bioprocessing.

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.

G. L. P. A. Ramos, J. S. Nascimento, L. P. Margalho, M. C. K. H. Duarte, E. A. Esmerino, M. Q. Freitas, A. G. Cruz, and A. S. Sant'Ana. Quantitative microbiological risk assessment in dairy products: Concepts and applications. *Trends in Food Science & Technology*, 111:610–616, 2021. ISSN 0924-2244.

C. R. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

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.

E. Robinson, M. Travanut, L. Fabre, S. Larréché, L. Ramelli, L. Pascal, A. Guinard, N. Vincent, C. Calba, L. Meurice, et al. Outbreak of salmonella newport associated with internationally distributed raw goats' milk cheese, france, 2018. *Epidemiology & Infection*, 148:e180, 2020.

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.

J. M. Ruzante, J. E. Lombard, B. Wagner, C. P. Fossler, J. S. Karns, J. A. S. Van Kessel, and I. A. Gardner. Factors associated with salmonella presence in environmental samples and bulk tank milk from us dairies. *Zoonoses and public health*, 57(7-8):e217–e225, 2010.

J. Sacks, S. B. Schiller, and W. J. Welch. Designs for computer experiments. *Technometrics*, 31(1):41–47, 1989.

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.

T. J. Santner, B. J. Williams, and W. Notz. *The design and analysis of computer experiments*. Springer, 2003.

N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger. Gaussian process optimization in the bandit setting: no regret and experimental design. In *In International Conference on Machine Learning*, 2010.

M. L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics. Springer New York, 1999.

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.

F. Tenenhaus-Aziza, J. J. Daudin, A. Maffre, and M. Sanaa. Risk-based approach for microbiological food safety management in the dairy industry: the case of listeria monocytogenes in soft cheese made from pasteurized milk. *Risk Analysis*, 34(1):56–74, 2014.

P. F. M. Teunis. Dose response for salmonella typhimurium and enteritidis and other nontyphoid enteric salmonellae. *Epidemics*, 41:100653, 2022.

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.

A. A. Tschuprow. On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron*, 2(3): 461–493, 1923.

A. Ung, A. Y. Baidjoe, D. Van Cauteren, N. Fawal, L. Fabre, C. Guerrisi, K. Danis, A. Morand, M. P. Donguy, E. Lucas, et al. Disentangling a complex nationwide salmonella dublin outbreak associated with raw-milk cheese consumption, france, 2015 to 2016. *Eurosurveillance*, 24(3):1700703, 2019.

H. Ungaretti Haberbeck, C. Plaza-Rodriguez, V. Desvignes, P. Dalgaard, M. Sanaa, L. Guillier, M. Nauta, and M. Filter. Harmonized terms, concepts and metadata for microbiological risk assessment models: The basis for knowledge integration and exchange. *Microbial Risk Analysis*, 10: 3–12, 2018. ISSN 2352-3522. doi: https://doi.org/10.1016/j.mran.2018.06.001. URL https://www.sciencedirect.com/science/article/pii/S2352352218300100. Special issue on 10th International Conference on Predictive Modelling in Food: Interdisciplinary Approaches and Decision-Making Tools in Microbial Risk Analysis.

J. A. S. Van Kessel, J. S. Karns, D. R. Wolfgang, E. Hovingh, and Y. H. Schukken. Dynamics of salmonella serotype shifts in an endemically infected dairy herd. *Foodborne pathogens and disease*, 9(4):319–324, 2012.

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 J. Bect. Sequential search based on kriging: convergence analysis of some algorithms. *arXiv preprint arXiv:1111.3866*, 2011.

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.

E. Vazquez, J. Villemonteix, M. Sidorkiewicz, and E. Walter. Global optimization based on noisy evaluations: an empirical study of two statistical approaches. In *Journal of Physics: Conference Series*, volume 135, page 012100. IOP Publishing, 2008.

J. Villemonteix, E. Vazquez, and E. Walter. An informational approach to the global optimization of expensive-to-evaluate functions, 2007.

Z. Wang and S. Jegelka. Max-value entropy search for efficient Bayesian optimization. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 3627–3635. PMLR, 06–11 Aug 2017.

W. J. Welch, R. J. Buck, J. Sacks, H. P. Wynn, T. J. Mitchell, and M. D. Morris. Screening, predicting, and computer experiments. *Technometrics*, 34(1):15–25, 1992.

World Health Organization. *Risk Management and Food Safety: Report of a Joint FAO/WHO Consultation, Rome, Italy, 27 to 31 January 1997*. Food & Agriculture Org., 1997.

World Health Organization. *Risk assessments of Salmonella in eggs and broiler chickens*, volume 2. Food & Agriculture Org., 2002.

World Health Organization et al. *WHO estimates of the global burden of foodborne diseases: foodborne disease burden epidemiology reference group 2007-2015*. World Health Organization, 2015.

World Health Organization et al. *Microbiological Risk Assessment–Guidance for food*, volume 36. Food & Agriculture Org., 2021.

T. Wright. The equivalence of neyman optimum allocation for sampling and equal proportions for apportioning the u.s. house of representatives. *The American Statistician*, 66(4):217–224, 2012.

T. Wright. Exact optimal sample allocation: More efficient than neyman. *Statistics and Probability Letters*, 129:50–57, 2017. ISSN 0167-7152.

Gang Zhao and Pirooz Vakili. Monotonicity and stratification. In *2008 Winter Simulation Conference*, pages 313–319, 2008.

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.